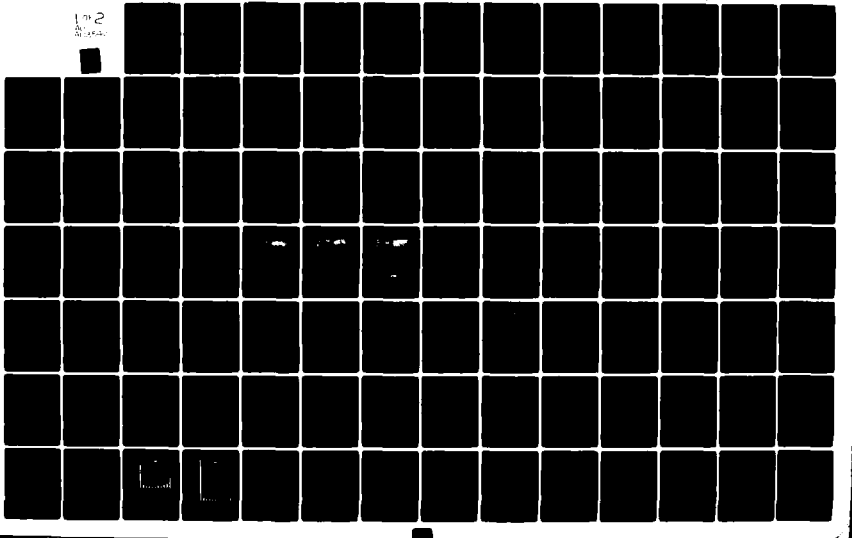
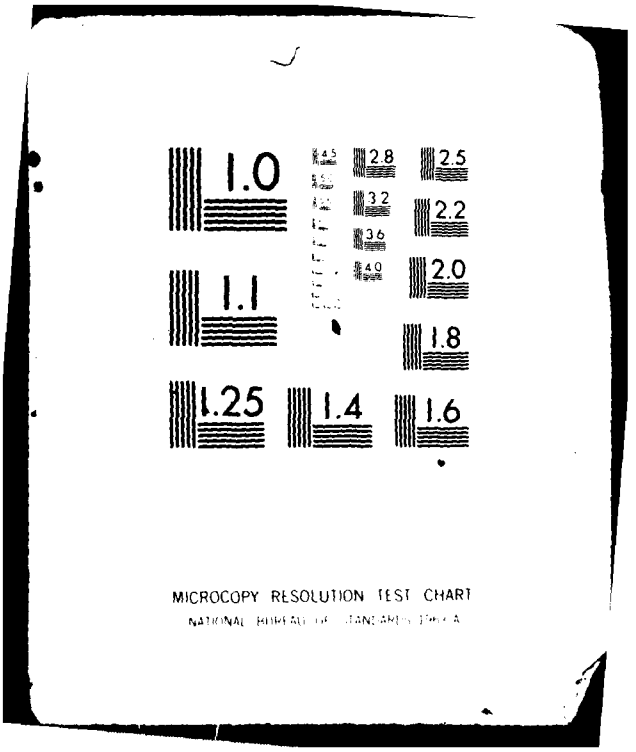
AD-A113 540 AIR FORCE ENVIRONMENTAL TECHNICAL APPLICATIONS CENTER--ETC F/G 4/2
CLOUD FORECAST SIMULATION MODEL.(U)
OCT 81 R C WHITON, E M BERECEK, J C SLADEN
UNCLASSIFIED USAFETAC/TN-81/004 SBI-AD-E850 149 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
MICROCOPY RESOLUTION TEST CHART

USAFETAC/TN-81/004

3

CLOUD FORECAST SIMULATION MODEL

by
Roger C. Whiton, Maj, USAF
Emil M. Berecek, Capt, USAF
John G. Sladen, 1st Lt, USAF

October 1981

UNITED STATES AIR FORCE
AIR WEATHER SERVICE (MAC)
USAF
ENVIRONMENTAL
TECHNICAL APPLICATIONS
CENTER
SCOTT AIR FORCE BASE, ILLINOIS 62225

REVIEW AND APPROVAL STATEMENT

USAFETAC/TN-81/004, Cloud Forecast Simulation Model, October 1981, is approved for public release. There is no objection to unlimited distribution of this document to the public at large, or by the Defense Technical Information Center (DTIC) to the National Technical Information Service (NTIS).

This technical publication has been reviewed and is approved for publication.

Peter J Havanac

PETER J. HAVANAC, Major, USAF
Chief, Aerospace Sciences Branch
Reviewing Officer

FOR THE COMMANDER

Walter S Burgmann

WALTER S. BURGMANN
Scientific and Technical Information
Officer (STINFO)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER USAFETAC/TN-81/004	2. GOVT ACCESSION NO. AD-A113540	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* CLOUD FORECAST SIMULATION MODEL		5. TYPE OF REPORT & PERIOD COVERED Technical Note 6. PERFORMING ORG. REPORT NUMBER USAFETAC Project 2339
7. AUTHOR*(s)* Roger C. Whiton, Maj, USAF Emil M. Berecek, Capt, USAF John G. Sladen, 1st Lt, USAF		8. CONTRACT OR GRANT NUMBER*(s)*
9. PERFORMING ORGANIZATION NAME AND ADDRESS US Air Force Environmental Technical Applications Center/DNS Scott AFB Illinois 62225		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS US Air Force Environmental Technical Applications Center Scott AFB Illinois 62225		12. REPORT DATE October 1981 13. NUMBER OF PAGES 134
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Climatology, clouds, cloud forecasting, cloud forecast simulation, computerized simulation, correlation, environment, environmental simulation, FCLDO (Cloud Forecast Simulation Model, Level 0), forecasting, forecast-observation correlation, forecast simulation, great circle distance, Gringorten Model-B, mathematical models, nephanalysis, normalization, sawtooth wave, scale (Cont'd)

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The cloud forecast simulation model generates synthetic worldwide 3-hour total cloud cover forecast fields at 50-nautical mile resolution. The synthetic forecasts are generated stochastically, based on input verifying "observed" total cloud cover fields, in such a manner that the agreement between the synthetic cloud forecast field and its verifying observed field is no better or worse than the agreement between actual cloud prognoses and their verifying observations. Moreover, a sawtooth wave submodel is used to insure the synthetic cloud forecast fields have the same spatial correlation as actual (Cont'd)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

19. KEY WORDS (Cont'd):

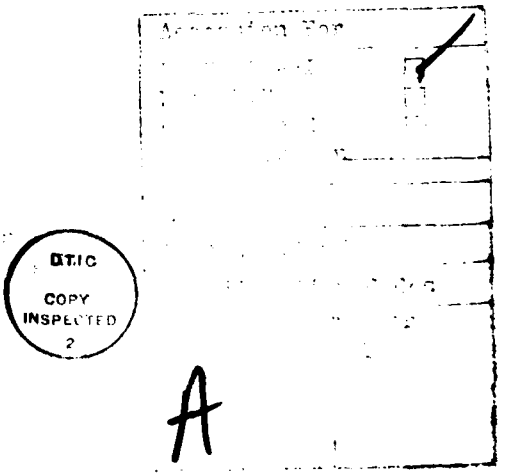
distance, simulation, skill, skill matrix, spatial correlation, statistics, stochastic modeling, tetrachoric correlation, weather, meteorology.

20. ABSTRACT (Cont'd):

cloud prognoses have. Thus, the cloud forecast simulation model generates synthetic total cloud cover forecast fields that have the same skill and spatial correlation as the operational forecast product has. The model is used to generate meteorological input to system planning and optimization simulations and system design studies. The sawtooth wave submodel could also be used to generate synthetic two-dimensional observed weather fields as well as cloud forecasts.

ACKNOWLEDGMENT

The sawtooth wave model used to generate spatially correlated, two-dimensional fields of random normal numbers was originally developed by Major Albert R. Boehm, USAFETAC/DNP, for USAFETAC Project 1960, COLOSSUS Weather Simulation. Major Boehm's helpful technical advice during the early stages of the present project is gratefully acknowledged.



CONTENTS

LIST OF ILLUSTRATIONS

PAGE

LIST OF TABLES

Chapter 1

INTRODUCTION

For verily not by design did the first beginnings of things station themselves each in its right place guided by keen intelligence; nor did they bargain sooth to say which motions each should assume, but because many in number and shifting about in many ways throughout the universe they are driven and tormented by blows during infinite past. After trying motions and unions of every kind, at length they fall into arrangements such as those out of which this our sum of things has been formed.

--Lucretius
De Rerum Naturae, c. 98-55 B.C.

Given for one instant an intelligence that could comprehend all the forces by which nature is animated and the respective positions of the beings that compose it, if moreover this intelligence were vast enough to submit these data to analysis, it would embrace in the same formula both the movements of the largest bodies in the universe and those of the lightest atom. To it, nothing would be uncertain, and the future as the past would be present to its eyes.

--Laplace, Oeuvres, Vol VII, 1812-1820

In meteorology, the number of particles concerned is so enormous that an accurate record of their initial positions and velocities is utterly impossible; and if this record were actually made, and their future positions and velocities computed, we should have nothing but an impenetrable mass of figures which would need a radical reinterpretation before it could be of any service to us. The terms "cloud," "temperature," "turbulence," etc., are all terms referring not to one single physical situation but to a distribution of possible situations of which only one actual case is realized. If all the readings of all the meteorological stations on earth were simultaneously taken, they would not give a billionth part of the data necessary to characterize the actual state of the atmosphere from a Newtonian point of view. They would give only certain constants consistent with an infinity of different atmospheres, and at most, together with certain a priori assumptions, capable of giving as a probability distribution, a measure, over the set of possible atmospheres. Using the Newtonian laws, or any other system of causal laws whatever, all we can predict at any future time is a probability distribution of the constants of the system, and even this predictability fades out with the increase of time.

--Norbert Wiener, Cybernetics, 1948

> Certain formally deterministic fluid systems possessing many scales of motion may be observationally indistinguishable from indeterministic systems, in that they possess an intrinsic finite range of predictability which cannot be lengthened by reducing the error of observation to any value greater than zero. Specifically, at any particular range there is a definite limit beyond which the expected accuracy of a prediction cannot be increased by reducing the uncertainty of the initial state to a fraction of its existing size. In this respect, these systems are like indeterministic systems, differing only in that the latter systems cannot be perfectly predicted even when the uncertainty of the initial state is reduced to zero.
>
> --E. N. Lorenz, Tellus, 21 (1969), 289-307

Short-range point forecasts of total cloud cover are made operationally by the Air Force Global Weather Central (AFGWC). For purposes of this technical note, the term "total cloud cover" indicates the fraction of the ground covered at any one time by clouds at all altitudes, as viewed vertically downward. These operational AFGWC cloud prognoses have an inherent spatial resolution of approximately 25 nautical miles (NM) (AFGWC 1/8-mesh grid) but are smoothed to 50-NM resolution (AFGWC 1/4-mesh grid) before release. The forecasts are made for 3 hours into the future and are valid at 12 Local Sun Time (LST). The meteorologist's term "valid" indicates the verifying time of the forecast. A forecast valid at 12 LST will be verified or "scored" against the actual or "observed" weather occurring at 12 LST.

Certain real time operational decisions are made in part on the basis of these operational AFGWC cloud prognoses. In order to conduct system design and optimization studies, it is useful to be able to simulate the entire process of real time, operational decision making in a nonreal time, nonoperational environment. Then, the expected effects of proposed changes in system mix and operating policy can be tested without the risk of harming real time operations

or adversely impacting the actual physical system. Simulating the entire operational system requires simulating the key components of the system, including in this case the AFGWC cloud prognoses. The generated synthetic or "simulated" forecasts of the total cloud cover must resemble those issued operationally by AFGWC.

Since approximately 1970, this need for simulated total cloud cover forecasts has been partially met by a technique developed by Detachment 1, HQ AWS. In the late 1970s, increasing sophistication in the system design and optimization studies being conducted made it apparent that the simple existing technique could no longer meet analysts' needs, and a search for an improved methodology began. As a first step in this search, the characteristics of a "good" cloud forecast simulation technique had to be considered. Clearly one such requirement is that the synthetic or simulated forecasts should be neither better nor worse than the actual forecasts.

All weather forecasts, including the AFGWC cloud prognoses, are imperfect to some degree. The "goodness" of weather forecasts is described in terms of their "skill." The skill of weather forecasts varies according to the type of forecast being made (e.g., tornado warnings are more difficult to make than cloud forecasts) and according to the location and time-of-year (because climatology exerts such a strong influence on the skill of forecasts). Moreover, for a given type of forecast at a given place and time, skill depends strongly on the duration of the forecast (e.g., it is less difficult to make a 3-hour cloud cover forecast than a 24-hour forecast of the same type). A long run of simulated forecasts of a given type and duration at a given place and time should not be significantly more skillful than a long run of operational

forecasts of the same type and duration for the same place and time. But having about the same skill as the operational forecasts is not the only requirement to be met by simulated forecasts. The spatial correlation of the simulated forecasts should be similar to that of the actual forecasts, whose spatial correlation, in turn, should not differ greatly from that of weather observations of the same type.

Just as the weather at one location is not independent of that at other, nearby locations, so forecasts of the weather at one location should be correlated spatially with forecasts at other, not too distant locations. The forecasts for two points separated by a distance of zero should be in perfect agreement, i.e., have a correlation of unity. As the distance between the locations increases, the correlation between the forecasts should decrease. At some large distance separating the two locations the weather at one may be independent of that at the other; at that distance it is reasonable to assume that the forecasts are also independent, i.e., have a correlation of zero. Thus, the spatial correlation of weather observations and forecasts decreases with increasing distance. A good forecast simulation technique should produce synthetic forecasts that are correlated in space in much the same manner as the actual forecasts (and observations).

The cloud forecast simulation technique currently employed by Detachment 1, HQ AWS meets the first requirement, i.e., it produces synthetic forecasts which, in the long run, have the same skill as the operational forecasts. Their technique fails to preserve spatial correlation, however. Despite efforts to operate on the grid system in a groupwise fashion (5 x 5 grid points

at a time), the technique produces synthetic forecasts having a "salt and pepper" appearance, lacking appropriate correlation in space.

In March 1980, USAFETAC was asked to develop a cloud forecast simulation technique capable of replacing the existing method and preserving both skill and spatial correlation in the synthetic cloud forecast fields. This was the genesis of a 1-year technique development effort during which the present authors developed, tested and delivered a successful prototype cloud forecast simulation model (FCLDO) that fully met the skill and spatial correlation requirements. The customer, after completing his own testing, accepted USAFETAC's Cloud Forecast Simulation Model and implemented it operationally. As a side benefit, the technique development software produced for this project --consisting of FORTRAN computer programs FCLDO and FCLDJ--were also delivered to the customer in various versions throughout the duration of the project. Toward the end of the project, key elements of this software were incorporated by the customer in his operational simulation software.

This work has application beyond the scope of of the original project. In order to generate the spatially correlated synthetic forecasts, a two-dimensional sawtooth wave field simulation submodel--originally developed by Major Albert R. Boehm, USAFETAC--was used to produce spatially correlated fields of random normal numbers. Using the sawtooth wave submodel in this way served to extend the model development capabilities to two spatial dimensions plus time.

The remaining chapters of this technical note detail the requirements imposed on the Cloud Forecast Simulation Model, explain the model itself, list its important assumptions and limitations, and describe its performance in terms of synthetic skill and spatial correlation.

Chapter 2

CLOUD FORECAST SIMULATION REQUIREMENTS

The basic requirement calls for development of a Cloud Forecast Simulation Model capable of generating simulated or synthetic forecasts of the total cloud cover with desired skill and spatial correlation.

In particular, the model must simulate AFGWC's operational point forecasts of the total cloud cover. These forecasts are produced at 50-NM resolution (AFGWC 1/4-mesh grid) by applying a 9-point, 4-2-1 smoother to the worldwide, 25-NM resolution (AFGWC 1/8-mesh grid) SAVDOX cloud prognosis data base. The SAVDOX data base is actually a composite of the output of three AFGWC cloud prognosis models: TRONEW, 5LYR, and HRCP (described in AFGWC publications). This composite SAVDOX data base consists of 3-hour forecasts of the total cloud cover, valid at 12 LST. The simulation model, therefore, must produce a two-dimensional, 50-NM resolution, worldwide field or network of synthetic 3-hour total cloud cover forecasts valid at 12 LST. The number of points requiring simulated forecasts was sufficiently large to call for use of two-dimensional field simulation techniques rather than methods such as USAFETAC's Multivariate Triangular Matrix Model. It is suitable for simulating weather at a small number of locations, but is too cumbersome to apply to problems involving more than 10-15 points.

Synthetic total cloud cover forecasts produced by the Cloud Forecast Simulation Model must not be significantly more skillful or less skillful than the

operational AFGWC cloud prognoses. The skill of the operational product is measured in terms of monthly, hemispheric, 21 x 21 verification contingency tables, called "skill matrices" or "phi matrices." The skill matrices are described in detail in Section 3.6 of this technical note. Basically, element S(i,j) of a raw count skill matrix $\underline{S}$ would contain a count of the number of times that a forecast cloud coverage category j later verified as observed cloud coverage category i. Synthetic skill matrices generated by a long run of the simulation model should converge to the operational AFGWC skill matrices for the month and hemisphere being modeled, i.e., the simulation must preserve the AFGWC skill matrices.

The simulated total cloud cover forecast fields, moreover, must not have the random, salt-and-pepper appearance that would result from simulating the forecast for each grid point independently from the forecast for neighboring grid points. Instead, the synthetic forecast fields must have a realistic spatial correlation structure not unlike that of the AFGWC SAVDOX forecast fields, smoothed to 50-NM resolution. Assuming spatial correlation is isotropic permits using a directionally independent spatial correlation function of distance d,

$$\rho = \rho(d) \tag{1}$$

as the measure to be preserved in the simulation. Accordingly, the spatial correlation function of synthetic cloud forecast fields produced by the model must not differ significantly from that of smoothed SAVDOX forecast fields. That requirement proved to be a complicating factor in model development because AFGWC does not archive the SAVDOX fields, nor has anyone studied the

spatial correlation of the SAVDOX fields. Since SAVDOX fields are 3-hour forecasts made from 3DNEPH with simple advection dominating in the short range, smoothed SAVDOX should have a spatial correlation function much like that of smoothed 3DNEPH data. Following this reasoning, the spatial correlation function of the model output fields was adjusted to match the spatial correlation function of smoothed 3DNEPH data. The latter were especially studied for that purpose. This is what is meant by the shorthand statement, the simulation must preserve spatial correlation of forecasts.

The final requirement has to do with the manner in which the simulation model must operate. The model must generate simulated 12 LST forecasts from input 12 LST verifying Multi-purpose Simulator (MPS) total cloud cover fields. MPS is a special cloud cover data base derived from time-sliced 3DNEPH data and smoothed to 50-NM (AFGWC 1/4-mesh grid) resolution. Input 3DNEPH data are available for each of the eight synoptic hours (00, 03, 06, 12, 15, 18, and 21 GMT) for each day, for grid points 25-NM apart (AFGWC 1/8-mesh grid). Time slicing involves dividing the hemisphere into eight slices or wedges of 45 degrees longitude each. The LST-to-GMT correction of the center of each such wedge is then taken as the correction factor for all grid points in the entire wedge. Thus, fields of cloud cover data at a roughly "constant" local sun time are produced by applying this time correction factor or "time slicing" to the input "constant GMT" 3DNEPH data. After being time sliced to produce cloud cover fields at a "constant" LST, the data are then smoothed by a 9-point, 4-2-1 smoother to produce the 50-NM resolution (AFGWC 1/4-mesh grid) MPS data. These input 12-LST MPS data are actually the verifying data for the simulated 12-LST forecasts, not the observed data on which the operational forecasts would, in real time, have been based. The insistence on having the model gen-

erate its synthetic forecasts from input verifying MPS data had the effect of specifying the basic design of the model. It was decided to generate the synthetic forecasts by decorrelating them from the verifying MPS data, with the amount of decorrelation being proportional to the skill of the forecasts.

In summary, the requirements to be met by the Cloud Forecast Simulation Model are as follows:

a. Simulate AFGWC SAVDOX total cloud cover forecasts
   (1) 3-hour forecasts
   (2) 50-NM (AFGWC 1/4-mesh grid) resolution
   (3) Worldwide
   (4) Valid at 12 LST

b. Preserve forecast skill matrices

c. Preserve spatial correlation of forecasts

d. Generate simulated 12-LST forecasts from input 12 LST verifying MPS fields

Chapters 3 and 6 of this technical note describe how USAFETAC's Cloud Forecast Simulation Model (FCLDO) meets requirements a and d. Chapters 3 and 4 describe how the model meets requirement b. Chapters 3 and 5 describe how the model meets requirement c.

Chapter 3

CLOUD FORECAST SIMULATION MODEL

3.1 Basic Model.

Let $\underline{c}_o$ be a field of observed cloud cover at time t, and let $\underline{\overset{''}{c}}_o$ be the corresponding field of equivalent normal deviates (ENDs)* of the observed cloud cover at time t. Let $\underline{c}_f$ be the field of cloud cover forecasts to occur at time t, and let $\underline{\overset{''}{c}}_f$ be the field of ENDs of the forecast cloud cover. Thus, $\underline{c}_o$ and $\underline{\overset{''}{c}}_o$ are the verifying fields for $\underline{c}_f$ and $\underline{\overset{''}{c}}_f$. Naturally the forecast and observed fields both refer to time t, even though the forecasts may have been based on another observed field at time $t - \Delta t$.

The extent to which a forecast valid at time t verifies can be expressed in terms of the correlation between the forecast field at time t and the verifying observed field at t. If the forecast and observed fields agree perfectly, their correlation can be expected to be unity. If at time t_o a forecast is made for the same time, t_o, that forecast can be expected to agree perfectly (have a correlation of unity) with the observation at time t_o. If the forecast is made for a time,

$$t = t_o + \Delta t \tag{2}$$

*Equivalent normal deviates or ENDs, are explained in USAFETAC/DNS ltr, 11 Apr 80, Status Report No. 2, USAFETAC Project 2082, TALON Weather Simulation Model, and in AWS-TR-75-259.

in the future, then the agreement between the forecast and verifying observation is almost always less than perfect. In general, the forecast observation correlation ρ_{fo} decreases as the forecast time step Δt lengthens. Often the data support an exponential decay of ρ_{fo} such as

$$\rho_{fo} = 0.96^{\Delta t} \qquad \text{or} \qquad \rho_{fo} = 0.98^{\Delta t} \tag{3}$$

where Δt is the forecast length in hours. In such models, when $\Delta t = 0$ hours, $\rho_{fo} = 1$; when $\Delta t = 1$ hour, ρ_{fo} 0.96 or 0.98. The more skillful the forecasting technique in use, the larger the constant in these forecast-observation correlation decay models.

Let $\underline{\rho}_{fo}$ represent the correlation between equivalent normal deviates $\underline{\overset{''}{c}}_o$ of the observed total cloud cover at time t and the corresponding ENDs $\underline{\overset{''}{c}}_f$ of the forecast total cloud cover at t. Conceivably, a forecast-observation correlation value could be provided for every position in the fields of $\underline{\overset{''}{c}}_o$ and $\underline{\overset{''}{c}}_f$; this is why $\underline{\rho}_{fo}$ is shown as a matrix. From meteorological considerations, it is reasonable to expect that variations in the forecast-observation correlation (i.e., variations in forecast skill) will be at a scale greater than that characteristic of the field being forecast. In practice, one is hard pressed to justify statistically the claim that adjacent values of ρ_{fo} are significantly different from one another. As a first approximation, this model will use a single, scalar value ρ_{fo} to represent the field of forecast-observation correlations $\underline{\rho}_{fo}$. The forecast field $\underline{\overset{''}{c}}_f$ can be simulated to preserve the forecast-observation correlation ρ_{fo} by a stochastic model such as that in Figure 1.

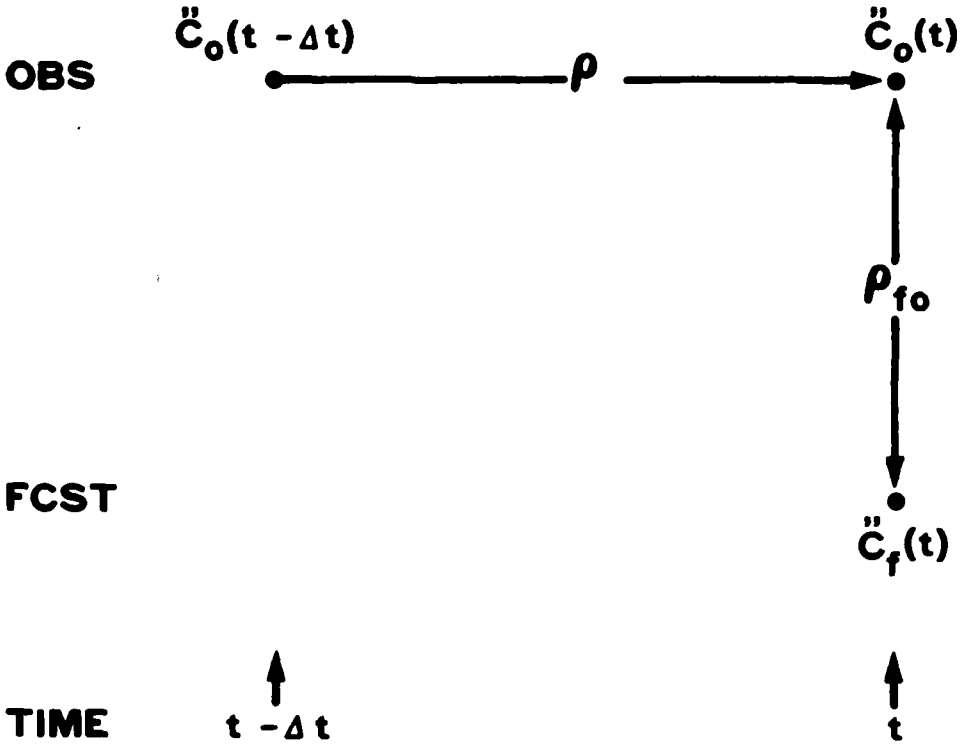


Figure 1. Model for Simulation of Forecast Fields in Such a Way as to Preserve Correlation Between Forecast Valid at Time t and Observation at Time t.

In the model shown in Figure 1, the simulated forecast field at time t depends strictly on the observed, or "verifying" field at the same time t. The correlation between forecast and observations at time t is preserved. No effort is made in this model to preserve other correlations, such as the correlation between the forecast field at time t and the observed field at time t - Δt on which the forecast was based.

The overall plan of the model is shown in Figure 2. The user of the model provides, for time t, a field $\underline{c}_o$ of total cloud cover. The model converts this

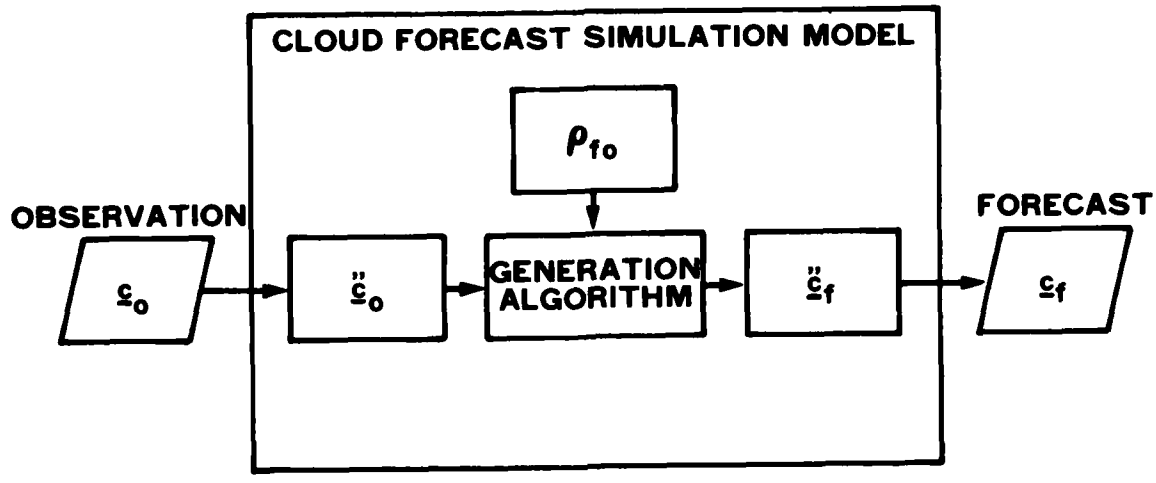


Figure 2. Macro-Design of Cloud Forecast Simulation Model Showing Flow of Information Through the Model. User provides observation field at time t. Model generates simulated forecast field at time t. The entire process is run again at time t + Δt, with the user providing an observation field for t + Δt and the model generating a simulated forecast field for t + Δt.

field to a field $\underline{\underline{c}}_o''$ of equivalent normal deviates of total cloud cover. On the basis of that field and a previously established correlation ρ_{fo}, the forecast simulation model generates, for time t, a simulated forecast field $\underline{\underline{c}}_f''$ of ENDs that is statistically "consistent" with the observed verifying field at the same time. The model then converts the forecast END field $\underline{\underline{c}}_f''$ to a simulated forecast total cloud cover field $\underline{c}_f$ for time t. At the succeeding time, t + Δt, the user provides another observation field $\underline{c}_o$, and the whole process is repeated.

The forecast simulation equation used in this model is

$$\underline{\underline{c}}_f = \rho_{fo}\underline{\underline{c}}_o + \sqrt{1 - \rho_{fo}^2}\,\underline{\eta} \qquad (4)$$

(a) (b)

The simulated forecast field is the sum of two components: (a) a deterministic part arising from the observed field $\underline{\underline{c}}_o$, and (b) a random or stochastic part allowing for the imperfect quality of weather forecasts. The random part of Equation (4) essentially decorrelates the forecast field from the observed, verifying field. In the case of perfect correlation between a forecast field $\underline{\underline{c}}_f$ and its verifying observation $\underline{\underline{c}}_o$, $\rho_{fo} = 1$, and the forecast field is determined completely by the observed field, i.e., the random part of Equation (4) plays no role whatsoever. In the case where there is no correlation between a forecast and its verifying observation (i.e., the forecast has no skill) $\rho_{fo} = 0$, and the random part of Equation (4) is the only contributor to the simulated forecast. In the intermediate case, where forecasts have some skill but are not perfect, $0 < \rho_{fo} < 1$, and a proportion of random error is added to the verifying field to generate the simulated forecast field. The amount of randomness added is governed by the forecast-observation correlation ρ_{fo}. For forecasts with a short lead time, ρ_{fo} will be large, and the random part of Equation (4) will contribute only weakly to the simulated forecast fields. For forecasts whose lead time is long relative to the predictability of the phenomenon modeled, ρ_{fo} will be small, and the random part of the solution will contribute strongly to the simulated forecast field.

The random field $\underline{\eta}$, a portion of which acts as the random part of Equation (4), must have a spatial structure or spatial correlation. This is because the

simulated forecast field $\underline{c}''_f$ is expected to have a spatial correlation approximating that of actual forecasts. If a field $\sqrt{1 - \rho_{fo}^2}\ \underline{\eta}$ that is uncorrelated in space were to be added to the spatially correlated $\rho_{fo}\underline{c}''_o$ field, the resulting $\underline{c}''_f$ field would show a spatial correlation substantially less than that of the input field $\underline{c}''_o$, and <u>much</u> less than that desired for the forecast field $\underline{c}''_f$ (in general, forecasts show a spatial correlation somewhat greater than that of observation fields). Under these circumstances, it becomes necessary to generate a field $\underline{\eta}$ of random normal numbers with desired spatial correlation. While this can be done by a variety of techniques, the one most suited to problems involving a large number of locations or grid points is the sawtooth wave method of Boehm (1979).

3.2 <u>Sawtooth Wave Submodel</u>.

The sawtooth wave submodel generates a field of equivalent normal deviates $\underline{\eta}$ having a desired spatial correlation function r.

Consider the correlation r between values η_j and $\eta_{j + \Delta t}$ located one grid distance Δj apart. This is shown in Figure 3.

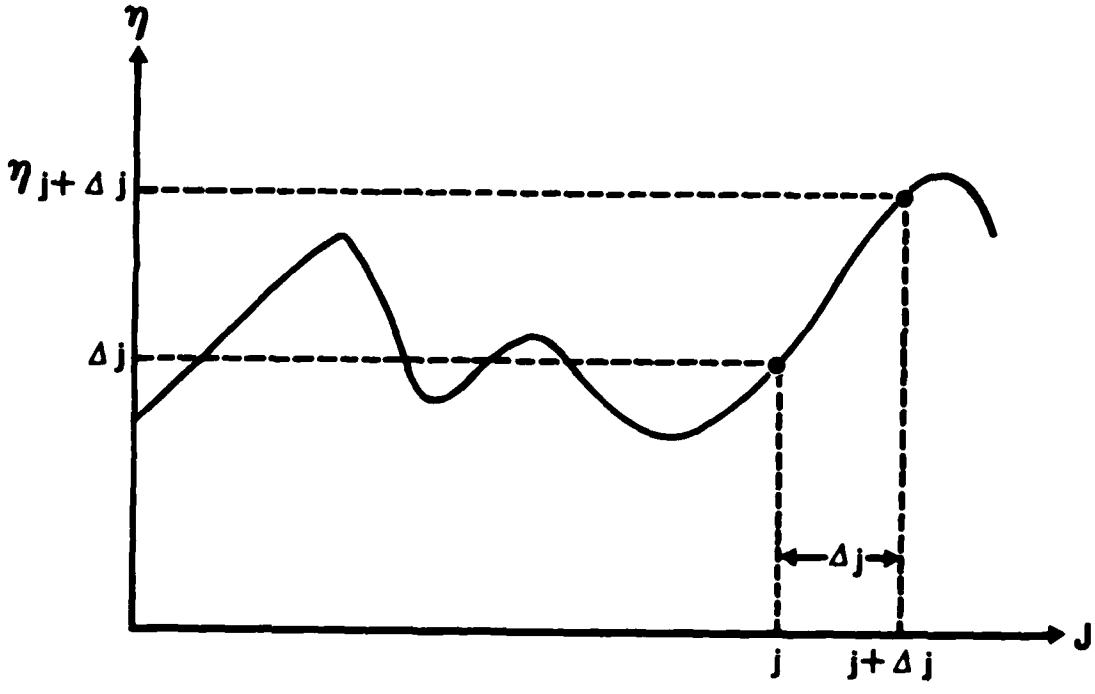


Figure 3. Correlation Between η at Location J = j and η at Location J = j + Δj Situated One Grid Distance Δj Apart.

Repeated samplings of the value of η at j and at j + Δj would produce a history of N data pairs from which the spatial correlation could be estimated by the Pearson product moment formula,

$$r = \frac{\frac{1}{N}\sum_{k=1}^{N} \eta_{j,k}\,\eta_{j+\Delta j,k} - \overline{\eta_j}\;\overline{\eta_{j+\Delta j}}}{s_{\eta_j}\;\;s_{\eta_{j+\Delta j}}} \qquad (5)$$

or by some other method. In Equation (5), the overbars represent means, and s represents the standard deviation. Since the η are ENDs, they are distributed normally with a mean of zero and a variance of unity. Therefore, for normally distributed η, Equation (5) reduces to

$$r = \frac{1}{N} \sum_{k=1}^{N} \eta_{j,k} \, \eta_{j+\Delta j,k} \tag{6}$$

Spatial correlation is being dealt with here. The correlation between η-values will be perfect (unity) at zero separation ($\Delta j = 0$) and will be less than or equal to unity with increasing distance Δj. To model the weather, a correlation function is needed that starts at unity and decreases with increasing distance d.

One such model that has been used successfully in ceiling, visibility and sky cover modeling is that of Gringorten (1979). In Gringorten's Model-B, the correlation function r depends on the geometric distance d and a characteristic scale distance D defined as the distance at which the correlation r falls to 0.99. In Gringorten's Model-B,

$$r = r(d,D) = \frac{2}{\pi} \left(\cos^{-1}\sigma - \sqrt{1 - \sigma^2} \right) \qquad \text{(dimensionless)} \tag{7}$$

$$\sigma = d/(128\ D) \qquad \text{(dimensionless)} \tag{8}$$

Because $\sigma \geq 0$, we can use the trigonometric relationship between arc cosine and arc sine to write,

$$r = r(d,D) \approx \frac{2}{\pi} (\sin^{-1}H - \sigma H) \tag{9}$$

$$H = \sqrt{1 - \sigma^2} = \sqrt{1 - d^2/(16384\ D^2)} \tag{10}$$

In this correlation function model, when the distance d equals the scale distance D, $\sigma = 1/128$, $H = 0.99997 \doteq 1$, $\sin^{-1}(1) = \pi/2$, $\sigma H = 0.00781$, and $r = 0.99$. Note that when $\sigma = 1$, $H \approx 0$, and $r = 0$. Therefore, Gringorten's Model-B correlation drops to zero at a distance $d = 128D$. Gringorten has estimated the scale distance for sky cover in Germany as 4 km. Using this for D in Equations (9) and (10) gives the correlation function shown in Figure 4. With this scale distance, the correlation drops to 0.99 in 4 km (2 NM) and to approximately zero at 512 km (276 NM).

It is desired that the sawtooth wave model produce a field of ENDs having the spatial correlation function of Gringorten's Model-B, discussed above.

In the sawtooth wave model, N sawtooth waves are allowed to emanate circularly from N focal points. Each focal point is the source of exactly one wave. The location of each focal point is picked at random, and the wavelength of each wave is selected at random from a range of permissible wavelengths. The field of equivalent normal deviates $\underline{\eta}$ is simply the sum of N sawtooth wave amplitudes at each grid point, corrected by subtraction of a constant.

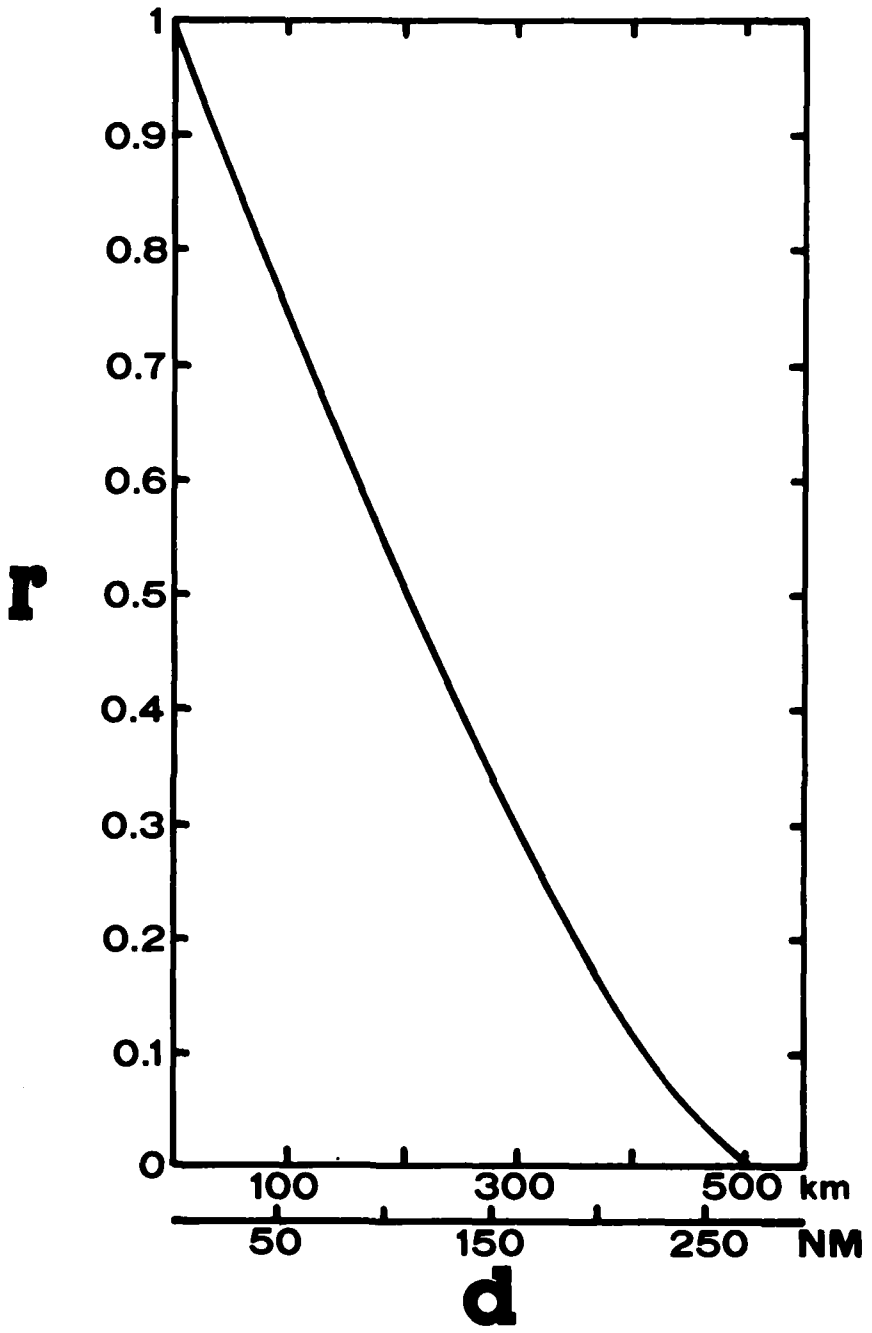


Figure 4. Correlation Function for Gringorten's Model-B with Scale Distance D = 4 km.

Each sawtooth wave is as shown in Figure 5. Amplitude of the wave, shown by y, varies between zero and one, depending on the observer's position along the wave. The sawtooth wave used here is a standing wave. Originating with zero amplitude at a focal point at distance d' = 0, it reaches maximum amplitude (unity) at distance d' = 1 wavelength, and thereafter falls to zero amplitude again. Within any one cycle of the sawtooth wave, the slope of wave amplitude versus distance is unity, i.e.,

$$dy/dd' = 1 \tag{11}$$

Hence, within any one cycle of the sawtooth wave, its equation is

$$y = d' \tag{12}$$

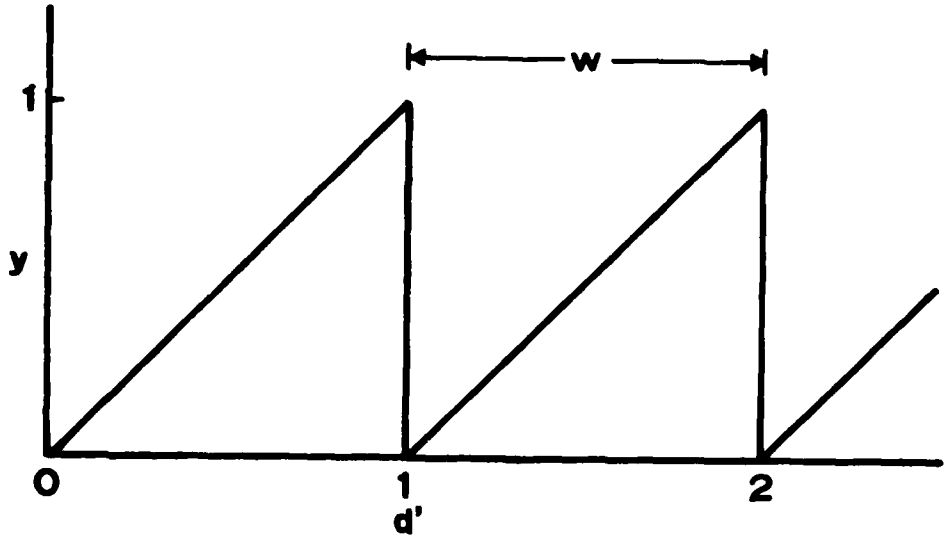


Figure 5. Sawtooth Wave. d' represents normalized distance orthogonal to the wave front. The normalized distance d' is measured in unit wavelengths, where wavelength is represented by w.

Allowing for multiple cycles of the sawtooth, Equation (12) becomes

$$y = d' - \mathrm{INT}(d') \tag{13}$$

where INT(d') represents the largest integer less than or equal to the normalized distance d'. The normalized distance d' is the geometric distance d expressed in unit wavelengths w, i.e.,

$$d' = d / w \tag{14}$$

Hence,

$$y = d/w - \mathrm{INT}(d/w) \tag{15}$$

An alternative Fourier representation of the sawtooth wave is

$$y = \pi - 2 \sum_{i=1}^{\infty} \frac{\sin id'}{i} \tag{16}$$

The simple form of the sawtooth wave makes it easy to calculate the amplitude y_{jk} of a wave at location j whose origin is the focal point at location k. This is done by computing the great circle distance between locations j and k, i.e.,

$$d = \mathrm{GCD}(j,k) \tag{17}$$

and then evaluating Equation (15) with a known wavelength w.

But any single wave amplitude y_{jk} does not create randomness. The $\underline{\eta}$ field produced by the sawtooth wave model must be random. Its elements η_j must have been selected at random from a normally distributed population with a mean of zero and a variance of one, i.e., N(0,1). The distribution of any one sawtooth wave is uniform, with a mean of 1/2 and a variance of 1/12. But the sum of approximately 12 uniform random numbers, by the central limit theorem, approaches the normal distribution. Naylor, et al. (1966) give the equation for calculating a normally distributed pseudorandom number G from the sum of N uniform pseudorandom numbers U:

$$G = \sigma_G \sqrt{\frac{12}{N}} \left(\sum_{n=1}^{N} U_n - \frac{N}{2} \right) + \mu_G \tag{18}$$

where σ_G and μ_G are the desired standard deviation and mean, respectively, of G. For the special case where $\sigma_G = 1$ and $\mu_G = 0$, and where the number of uniform random numbers to be summed is N = 12, Equation (18) simplifies to

$$G = \sum_{n=1}^{12} U_n - 6 \tag{19}$$

Calculating a normally distributed value η_j for location j from the sum of N = 12 uniformly distributed sawtooth wave amplitudes y_{jk}, Equation (19) becomes

$$\eta_j = \sum_{k=1}^{12} y_{jk} - 6 \qquad (20)$$

Figure 6 illustrates the superposition of sawtooth waves. Two sawtooth waves emanate from randomly positioned focal points k = 1 and k = 2. These waves converge on location j with respective amplitudes y_{j1} and y_{j2}. The wavelengths w_k of the two waves are different to emphasize that those wavelengths were drawn at random uniformly from a range of possible wavelengths.

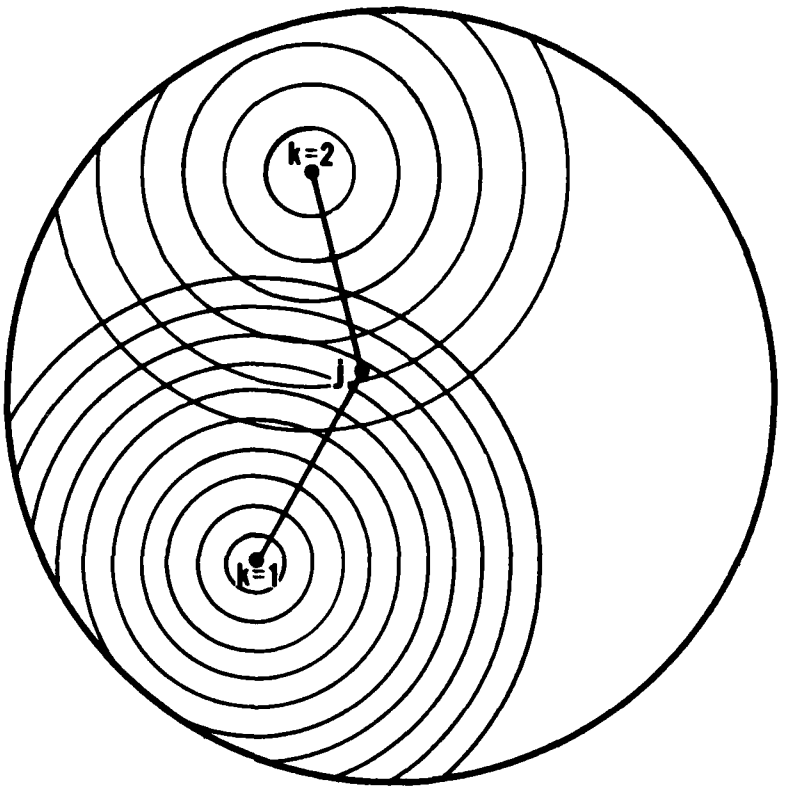


Figure 6. Sawtooth Emanating from Focal Points at Locations k Converge on Location j.

The great circle distance d between any two points "a" and "b" on the globe can be calculated from the latitude and longitude of point "a" (θ_a, λ_a) and that of point "b" (θ_b, λ_b). The conventional equation is

$$d = r \cos^{-1}[\sin\theta_a \sin\theta_b + \cos\theta_a \cos\theta_b \cos(\lambda_a - \lambda_b)] \tag{21}$$

where r is the radius of the earth, approximately 6371 km. This equation involves calculating five sines and cosines plus one arc cosine.

An alternate expression for the great circle distance d can be obtained by using the trigonometric function-product relations,

$$\sin\theta_a \sin\theta_b = (\tfrac{1}{2})[\cos(\theta_a - \theta_b) - \cos(\theta_a + \theta_b)] \tag{22}$$

$$\cos\theta_a \cos\theta_b = (\tfrac{1}{2})[\cos(\theta_a - \theta_b) + \cos(\theta_a + \theta_b)] \tag{23}$$

or

$$\sin\theta_a \sin\theta_b = (\tfrac{1}{2})(d - s) \tag{24}$$

$$\cos\theta_a \cos\theta_b = (\tfrac{1}{2})(d + s) \tag{25}$$

where

$$d = \cos(\theta_a - \theta_b) \tag{26}$$

$$s = \cos(\theta_a + \theta_b) \tag{27}$$

from which it is found that

$$d = r \cos^{-1}\{(\tfrac{1}{2})[(d - s) + (d + s)\cos(\lambda_a - \lambda_b)]\} \tag{28}$$

This equation requires three cosines and one arc cosine and should therefore be much faster to solve than Equation (21).

A third expression for the great circle distance can be obtained by using the trigonometric angle-difference relation,

$$\cos(\lambda_a - \lambda_b) = \cos\lambda_a \cos\lambda_b + \sin\lambda_a \sin\lambda_b \tag{29}$$

in Equation (21), from which it is found that

$$d = r \cos^{-1}[\sin\theta_a \sin\theta_b + \cos\theta_a \cos\theta_b(\cos\lambda_a \cos\lambda_b + \sin\lambda_a \sin\lambda_b)] \tag{30}$$

Because this equation involves eight sines and cosines plus one arc cosine, it appears at first glance much less suitable for use than Equations (21) or (28). Nevertheless, Equation (30) offers some "operational" advantages that make it useful. In particular, one need not know the actual latitudes and longitudes to calculate great circle distance from Equation (30); only the sines and cosines of the latitudes and longitudes are needed. Moreover, since the number of focal points is small (generally 12 or fewer), the needed sines and cosines can be calculated initially and then stored for repeated use.

Each sawtooth wave must emanate from a randomly positioned focal point; otherwise, the amplitude sums will not be random. Focal points are located in terms of their latitude θ_k and longitude λ_k, where, for convenience, the longitude ranges from 0 degrees through 360 degrees. The longitude λ_k of the kth focal point is selected uniformly from the range 0 degrees to 360 degrees by the equation,

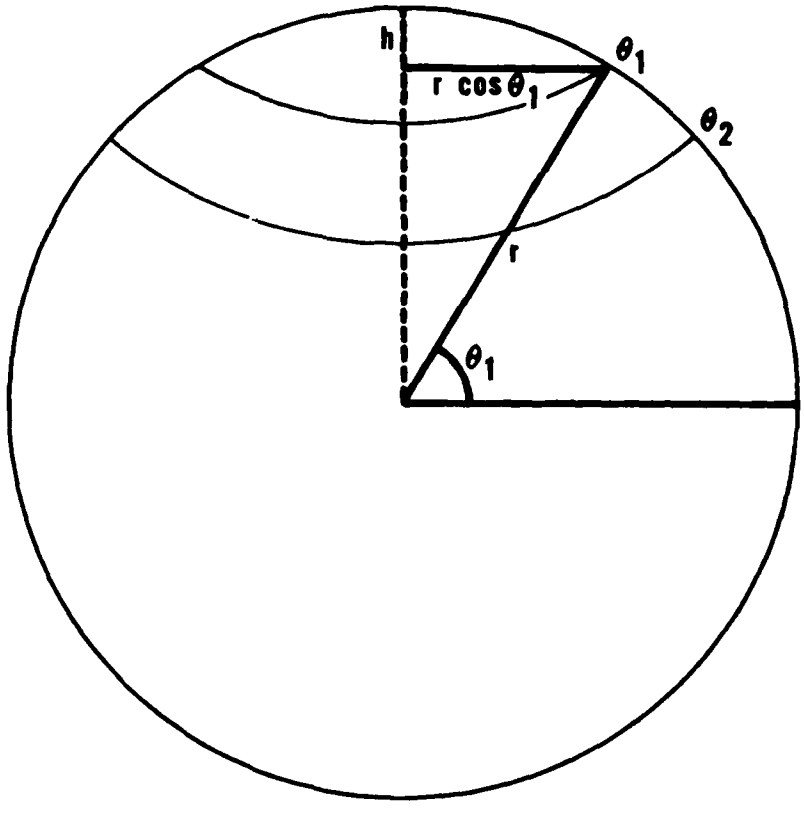
$$\lambda_k = 360\ U_k \tag{31}$$

where U_k is a pseudorandom number selected from a population uniformly distribution over the range (0,1).

While the longitude λ_k of the focal point can be selected uniformly from the range 0 degrees to 360 degrees, the latitude θ_k can not be selected uniformly from the range 0 degrees to 180 degrees (90 degrees to -90 degrees). This is because equiprobable latitude bands are not equal area bands, and simple selection of latitude would result in an overly dense concentration of focal points per unit surface area near the poles. Figure 7 shows the geometry of this problem. Needed is an expression for the surface area of the spherical zone bounded by latitudes θ_1 and θ_2. The essential principle is that the surface area of the zone is the difference between the surface area of the spherical cap formed by θ_2 and that formed by θ_1.

Let us consider only the spherical cap formed by θ_1. This has height h in a sphere of radius r. The surface area of that cap is

$$S_1 = 2\pi rh \tag{32}$$



igure 7. Geometry for Surface Area of the Spherical Zone Bounded by Latitudes θ_1 and θ_2, where $\theta_1 > \theta_2$.

But from the Pythagorean theorem,

$$x^2 = r^2 - r^2\cos^2\theta_1 = r^2(1 - \cos^2\theta_1) = r^2\sin^2\theta_1 \tag{33}$$

and

$$x = r\sin\theta_1 \tag{34}$$

Moreover,

$$h = r - x \tag{35}$$

$$h = r(1 - \sin\theta_1) \tag{36}$$

Hence, the surface area of the spherical cap formed by θ_1 is, from Equations (32) and (36),

$$S_1 = 2\pi r^2(1 - \sin\theta_1) \tag{37}$$

By analogy, the surface area of the spherical cap formed by θ_2 is

$$S_2 = 2\pi r^2(1 - \sin\theta_2) \tag{38}$$

The surface area S_z of the zone is the difference,

$$S_z = S_2 - S_1 \tag{39}$$

$$S_z = 2\pi r^2(\sin\theta_1 - \sin\theta_2) \tag{40}$$

The function difference relations give the result,

$$\sin\theta_1 - \sin\theta_2 = 2\cos\tfrac{1}{2}(\theta_1 + \theta_2)\sin\tfrac{1}{2}(\theta_1 - \theta_2) \tag{41}$$

Let us consider that the width of the latitude band θ_1 to θ_2 will always be constant, say 5 degrees or 10 degrees. Then the sine of one-half their difference is also a constant, say D:

$$\sin\tfrac{1}{2}(\theta_1 - \theta_2) = D \tag{42}$$

Thus,

$$\sin\theta_1 - \sin\theta_2 = 2\ D\ \cos\tfrac{1}{2}(\theta_1 + \theta_2) \tag{43}$$

Using Equation (43) in Equation (40) produces the result,

$$S_z = 4\pi D\ r^2\ \cos\tfrac{1}{2}(\theta_1 + \theta_2) \tag{44}$$

But r is constant, so

$$C = 4\pi D\ r^2 = \text{const} \tag{45}$$

and

$$\overline{\theta} = \tfrac{1}{2}(\theta_1 + \theta_2) \tag{46}$$

where $\overline{\theta}$ is the mean latitude of the zone bounded by θ_1 and θ_2. Using Equations (45) and (46) in Equation (44) produces the result,

$$S_z = C \cos \overline{\theta} \tag{47}$$

Equation (47) shows that the surface area of the spherical zone bounded by latitudes θ_1 and θ_2 is proportional to the cosine of the mean latitude of the zone. If we simply choose the latitude of the focal point uniformly over the permitted range of latitudes, then the density of selections will not show a poleward decrease proportional to the poleward decrease of zonal surface area

S_z. This can be compensated for by selecting cos θ_k rather than θ_k itself. Since the cosine has the range (0,1), the equation is

$$\cos \theta_k = U_k' \tag{48}$$

where U_k' is a uniform pseudorandom number drawn from the same range. Selection of the latitude of the focal point in this way restricts the focal point to the Northern Hemisphere, but imposes no limits on the randomness of the result.

If the wavelength w_k of the sawtooth wave emanating from location k is to be selected from the interval,

$$w_1 < w_k \leq w_2 \tag{49}$$

such that any value is equally likely to be chosen, then the selection can be made by drawing a random number uniformly from Equation (49). If U_k'' is a pseudorandom number drawn from a uniform distribution having the range 0 to 1, then

$$w_k = U_k''(w_2 - w_1) + w_1 \tag{50}$$

An algorithmic procedure for the sawtooth wave generator is shown in Figure 8. This is written loosely after the manner of ALGOL-68.

```
procedure  SAWTOO (m, ETA);
integer j, k, m, n;
real w, d, y
real array YSUM[1:m], ETA[1:m];
equivalence (YSUM, ETA);
for each location or grid point j: = 1 step 1 until m do
begin
   initialize YSUM_j: = 0.0;
end j;
for each focal point k: = 1 step 1 until n do
comment:  ... n = 12 ...;
begin
   select location of k^th focal point at random;
      comment:  ... Equations (31) and (48) ...;
   select wavelength w at random from (w_1,w_2);
      comment:  ... Equation (50) ...;
   for each location or grid point j: = 1 step 1 until m do
   begin
      calculate distance d: = GCD(j,k);
         comment:  ... Equation (30) ...;
      calculate wave amplitude y;
         comment:  ... Equation (15) ...;
      accumulate YSUM_j = YSUM_j + y;
   end j;
end k;
for each location or grid point j: = 1 step 1 until m do
begin
   ETA_j: = YSUM_j - 6;
      comment:  ... Equation (20) ...;
end j;
end SAWTOO;
```

Figure 8. Algorithm for Sawtooth Wave Submodel.

3.3 Normalization.

The transformation from raw total cloud cover to equivalent normal deviate of the total cloud cover, and vice-versa, is an important feature of the basic

model. This feature preserves the probability distribution of synthetic forecasts produced by the basic form of the model. A long run of this forecast simulation model will produce the same probability distribution of forecasts as that of the original data used to construct the model.

Selection of a normalization technique is necessarily a subjective procedure because it is influenced by the data being fitted. Based on the results of Somerville, Watkins, and Daley (1978) for sky cover, Johnson's S_B curve was used to fit the distributions of total cloud cover observations and forecasts. This curve is also called Johnson's bounded distribution and is described in Boehm (1976). In that distribution,

$$\overset{''}{c}_o = a + b \ln \left(\frac{c_o - c_{oL}}{c_{oU} - c_o}\right) \tag{51}$$

where $\overset{''}{c}_o$ is the END of the observed cloud cover c_o; c_{oL} is the lower bound of c_o, namely 0.0; and c_{oU} is the upper bound of c_o, namely 1.0. Thus,

$$\overset{''}{c}_o = a + b \ln \left(\frac{c_o}{1 - c_o}\right) \tag{52}$$

where a and b are coefficients that can be fitted by simple linear regression. By analogy, for cloud cover forecasts,

$$\overset{''}{c}_f = c + d \ln \left(\frac{c_f}{1 - c_f}\right) \tag{53}$$

Since verification contingency tables or skill matrices are available for each month, monthly Johnson curves were developed for both observations and forecasts. In other words, 12 each of Equations (52) and (53) were prepared from the margins of the verification contingency tables.

3.4 Model Input: Observed Data.

The field $\underline{c}_o$ of observed total cloud cover is taken from Multi-purpose Simulator data tapes for 12 LST. The MPS data are not true observations in that they are a time-sliced, smoothed form of 3DNEPH information, which in itself is a modeled composite of many, potentially conflicting data sources of differing type and scale.

3.5 Model Output: Synthetic Forecasts and Statistical Diagnostics.

The output available from the Cloud Forecast Simulation Model can generally be broken into two parts:

--displays of individual observed and forecast fields produced as the model is running

--reports displayed at the end of model execution describing the statistics of all observed and forecast fields produced by or input to the model

The output produced by any given run of the model is largely at the discretion of the user, as each of these products must be individually requested or rejected at the beginning of model execution.

The input observed and output synthetic forecast fields may be displayed on the printer in batch versions or written to a disk file in interactive version of the model. These outputs can be either data dump format, where the characters 0-9 and A-K represent cloud cover categories ranging from clear to cloudy (as displayed in Table 4), or in an analyzed display employing user-specified cloud cover categories and display symbols (see Figures 9-11). In an operational model, these fields could either be written to tape or used directly by another simulator.

The reports containing model performance analyses and statistics are mainly used for diagnostic purposes, and may be largely eliminated in an operational model. If requested, one report is generated for each month for which synthetic forecast fields were produced. The individual reports available, what information they contain, and a diagnostic guide for assessing model performance based on the contents of each report are displayed in Table 1.

3.6 Forecast Adjustment (Skill Matrix) Scheme.

The basic Cloud Forecast Simulation Model described above assumes a bivariate normal distribution of observations and forecasts. In actual fact, forecast verification contingency tables (called "skill matrices") for the type of forecast product being modeled here are not exactly bivariate normal. Under

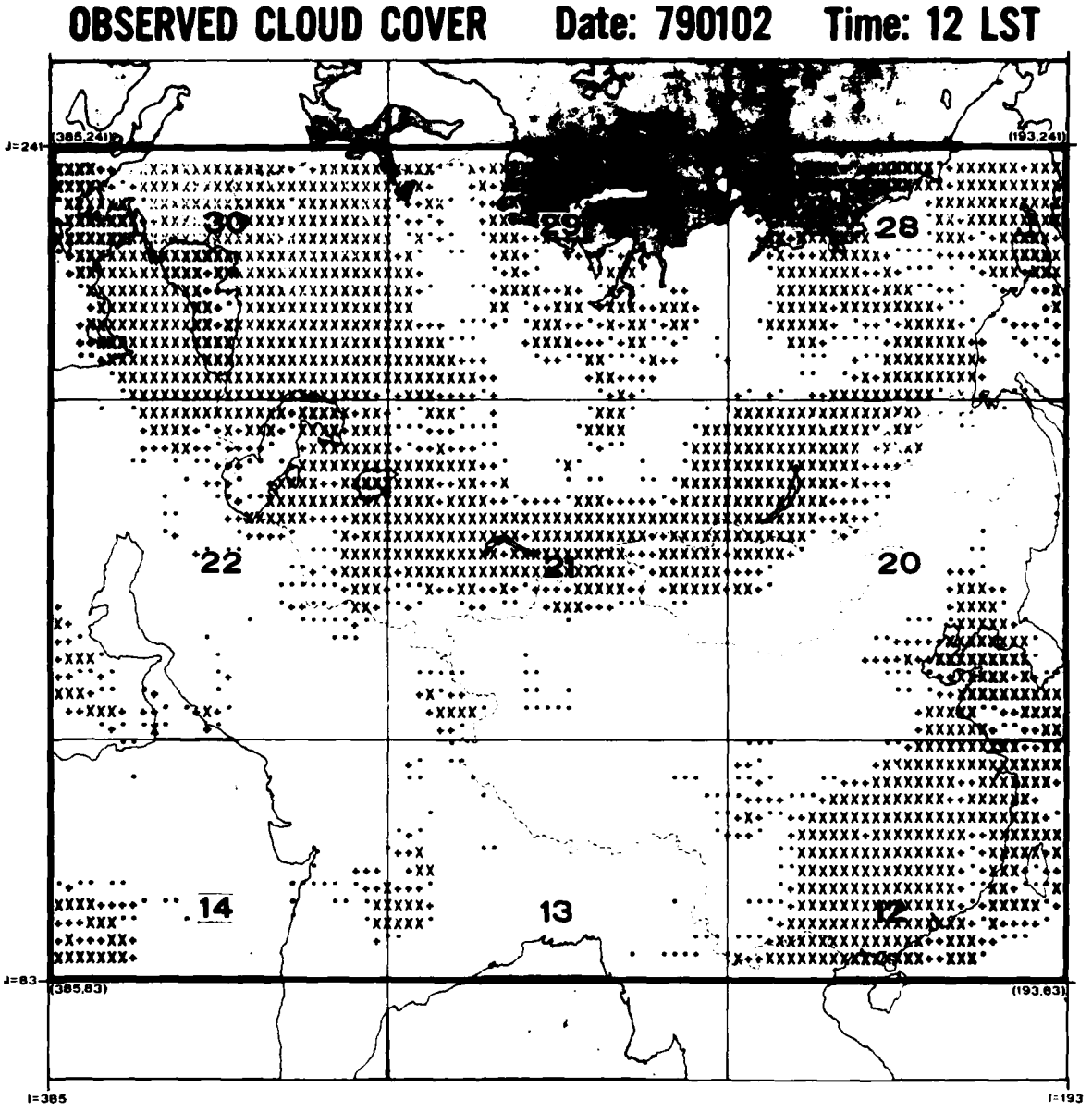


Figure 9. Observed Total Cloud Cover in 25-percent Coverage Categories from 2 January 1979, 12 LST MPS Data. The field of observed cloud cover is used as input to FCLDO to generate the fields of synthetic forecasts and then to verify the simulated forecast skill of the model.

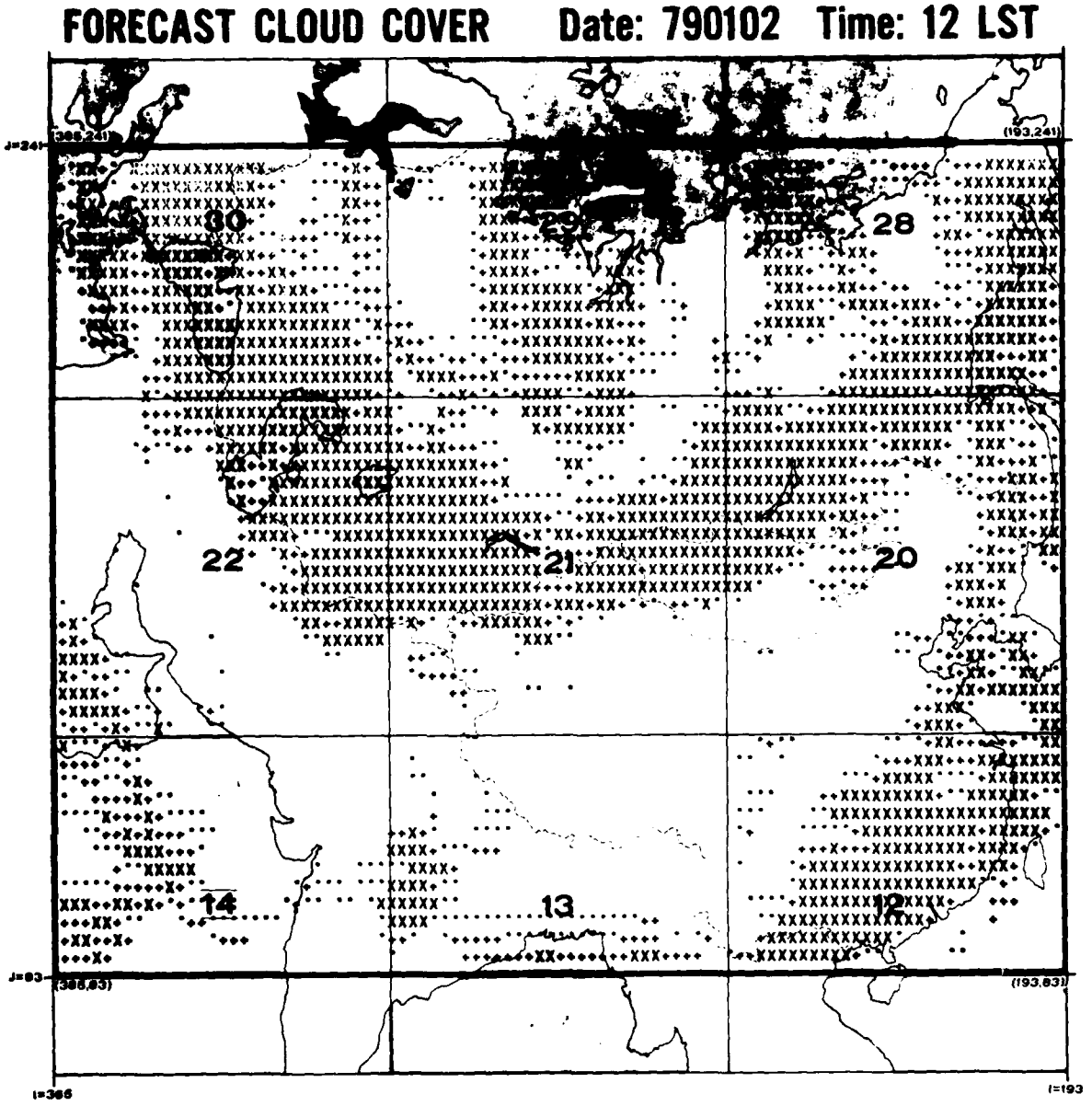


Figure 10. Simulated Field of Total Cloud Cover Forecasts in 25-percent Coverage Categories Valid 2 January 1979, 12 LST.

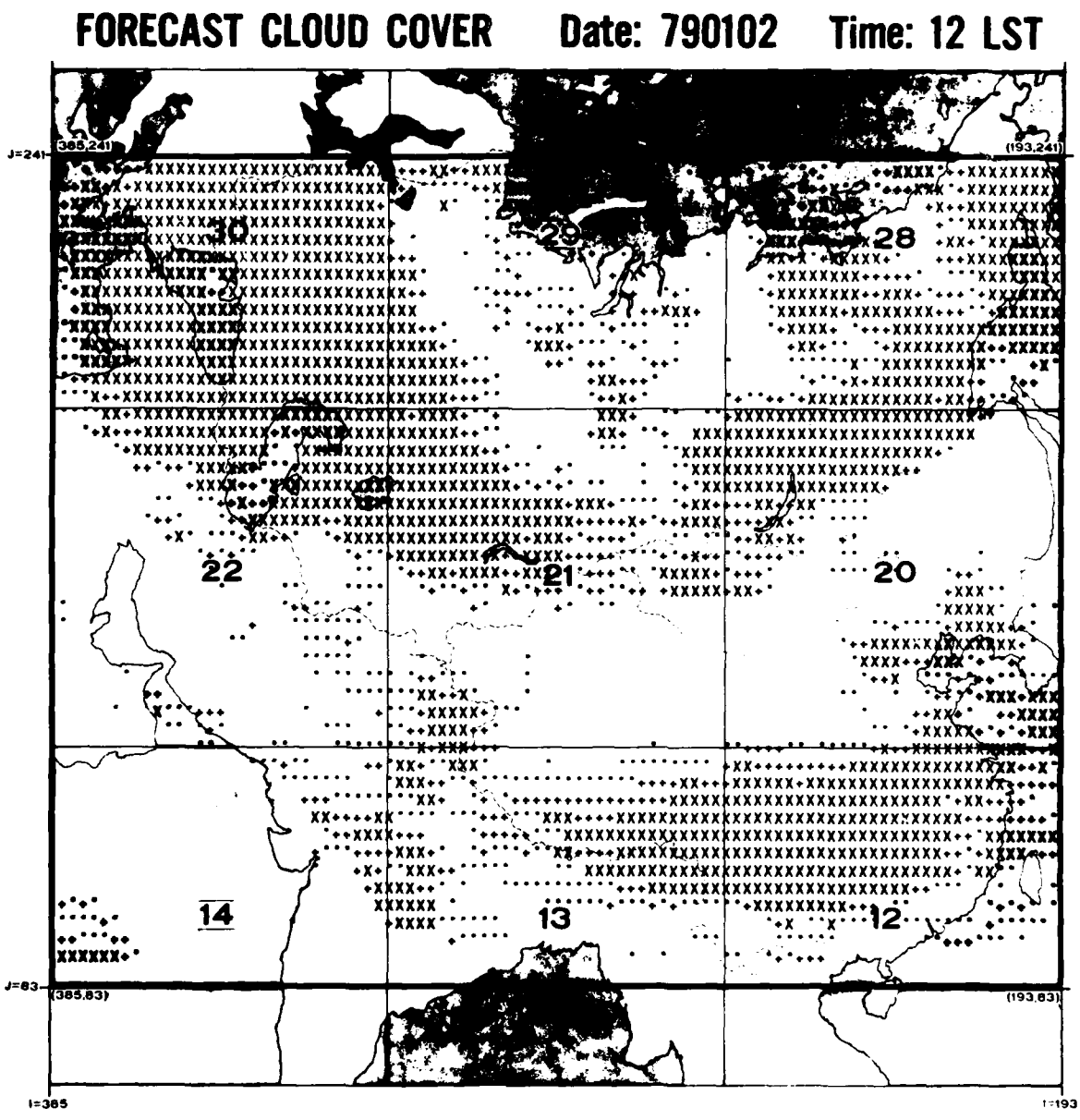


Figure 11. A Second Realization of Simulated Total Cloud Cover Forecasts in 25-percent Coverage Categories, Valid 2 January 1979, 12 LST. Although each realization was generated from the same input observation field and display the same spatial correlation characteristics and forecast vs. observation skill, they are not identical.

Table 1. Model Performance Reports and Diagnostic Information Available from the Cloud Forecast Simulation Model.

REPORT	CONTENTS	DIAGNOSTIC COMMENTS
Synthetic Skill Matrices		
-Raw Count Form	Raw count verification contingency table	----------
-Cumulative, Row-Normalized Percentage Form	Verification contingency table cumulated from clear to cloudy forecast categories for each observed category	Should compare favorably with AFGWC SAVDOX skill matrix for the same month in the same same form
Tetrachoric Forecast-Observation Correlation Information	Tetrachoric table and correlation computed from raw count synthetic skill matrix	Should compare favorably with forecast-observation correlation input if running the early version of FCLDO
Equivalent Normal Deviate Statistics	Mean observed and forecast cloud cover ENDs	Ideal value = 0
	Standard deviations of observed and forecast cloud cover ENDs.	Ideal value = 1
Skill Analysis of Synthetic Skill Matrix	Skill analysis and comparison of cumulative, row-normalized percentage form vs. AFGWC SAVDOX skill matrix in same form using a chi-square test.	Skill values should be similar for both matrices (especially for adjusted model described in Section 3.6 of of this report). Chi square value should be near zero (especially for the adjusted model).
Synthetic Skill Matrix Marginal Probability Distributions	Forecast and observed marginal probability distributions from synthetic skill matrix	----------
Spatial Correlations	Spatial correlations of model's forecast, observed and random (η) END fields	Spatial correlations for "ALL PTS" for all fields should be similar.

some circumstances--particularly when there is great confidence in the representativeness of the skill matrices--it may be desirable to make a statistical adjustment to the simulated cloud forecasts such that a long run of these synthetic forecasts will be distributed in accordance with the skill matrices.

Consider the January, Northern Hemisphere skill matrix (Table 2), and in particular the conditional distribution of forecasts given that the observed

Table 2. Skill Matrix for January, Northern Hemisphere.

OBSERVED	CLOUDY 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	CLEAR 21	TOTAL
1	27019	17871	8993	5507	3624	2524	1852	1399	1210	1065	876	790	705	655	566	563	584	424	386	343	498	77454
2	2291	2207	1475	1007	727	536	410	346	285	264	226	203	195	178	145	131	132	135	125	90	160	11268
3	2567	2717	1905	1511	1054	907	691	596	483	468	372	383	341	335	262	277	252	255	238	208	293	16115
4	802	894	721	580	452	413	313	288	218	195	194	166	166	168	131	143	116	119	115	131	146	6471
5	1025	1117	829	727	562	466	373	371	306	108	267	243	225	225	223	196	205	196	205	156	238	8463
6	656	714	559	502	445	350	303	267	228	262	230	201	190	178	180	161	182	165	151	152	239	6315
O 7	723	813	545	526	456	402	339	279	285	270	245	226	234	213	208	186	171	186	194	178	256	6935
B 8	493	512	409	286	278	238	229	221	197	172	177	141	158	153	117	124	143	128	126	121	174	4599
S 9	581	643	534	422	367	352	288	256	260	255	269	215	224	223	199	206	203	205	185	227	297	6411
E 10	255	241	192	176	176	139	132	137	146	116	129	130	117	138	137	102	127	113	116	130	197	3148
R 11	176	233	199	173	182	183	176	207	175	183	179	190	185	186	142	204	168	181	147	166	269	3924
V 12	105	121	91	101	108	104	91	94	96	113	110	98	125	123	103	103	116	101	103	117	206	2331
E 13	157	162	165	150	170	139	150	162	168	168	196	206	198	192	224	221	217	202	215	283	366	4111
D 14	69	92	77	67	86	63	88	96	97	82	95	94	117	135	103	128	121	131	131	153	238	2265
15	127	140	132	112	110	89	106	128	129	135	159	177	170	204	209	229	192	240	207	277	423	3695
16	105	116	106	121	105	107	73	107	116	98	145	144	174	148	171	201	223	243	263	292	503	3561
17	106	120	97	111	107	97	94	125	105	137	130	178	180	179	236	263	287	306	325	357	633	4173
18	70	85	55	50	72	81	63	75	87	85	94	104	121	144	172	208	224	268	292	303	566	3219
19	140	95	97	95	116	114	118	106	145	160	158	154	195	232	265	336	425	493	539	701	1477	6156
20	81	49	58	62	44	53	40	59	53	63	76	71	72	89	119	145	181	200	246	342	580	2683
21	747	587	469	424	463	501	464	550	602	623	684	769	940	1106	1445	1883	2528	3254	4219	6506	20830	49594
TOTAL	38295	29529	17708	12712	9704	7858	6393	5871	5393	5222	5011	4883	5032	5204	5354	6010	6797	7545	8528	11233	28611	232893

(verifying) total cloud cover is category 21 (clear). That row of the skill matrix is analyzed in Table 3 and illustrated in Figures 12 and 13.

From the relative frequencies in Figure 12, one can see that the conditional forecast distribution in the skill matrix is inverse J-shaped with a modal frequency near clear. Let us compare this conditional distribution from the skill matrix with the conditional forecast distribution generated by the model.

Consider the forecast simulation equation,

$$\underline{u}_f = \rho_{fo}\underline{u}_o + \sqrt{1 - \rho_{fo}^2}\,\underline{\eta} \tag{54}$$

which produces the spatially correlated forecast equivalent normal deviate (END) field $\underline{u}_f$. For any forecast grid point the field notation may be removed to yield

$$u_f = \rho_{fo}u_o + \sqrt{1 - \rho_{fo}^2}\,\eta \tag{55}$$

where η is now, in effect, a random normal END. For any given month, the forecast-observation correlation ρ_{fo} used by this model is a constant, and the simulation equation reduces to

$$u_f = G\,u_o + H\eta \tag{56}$$

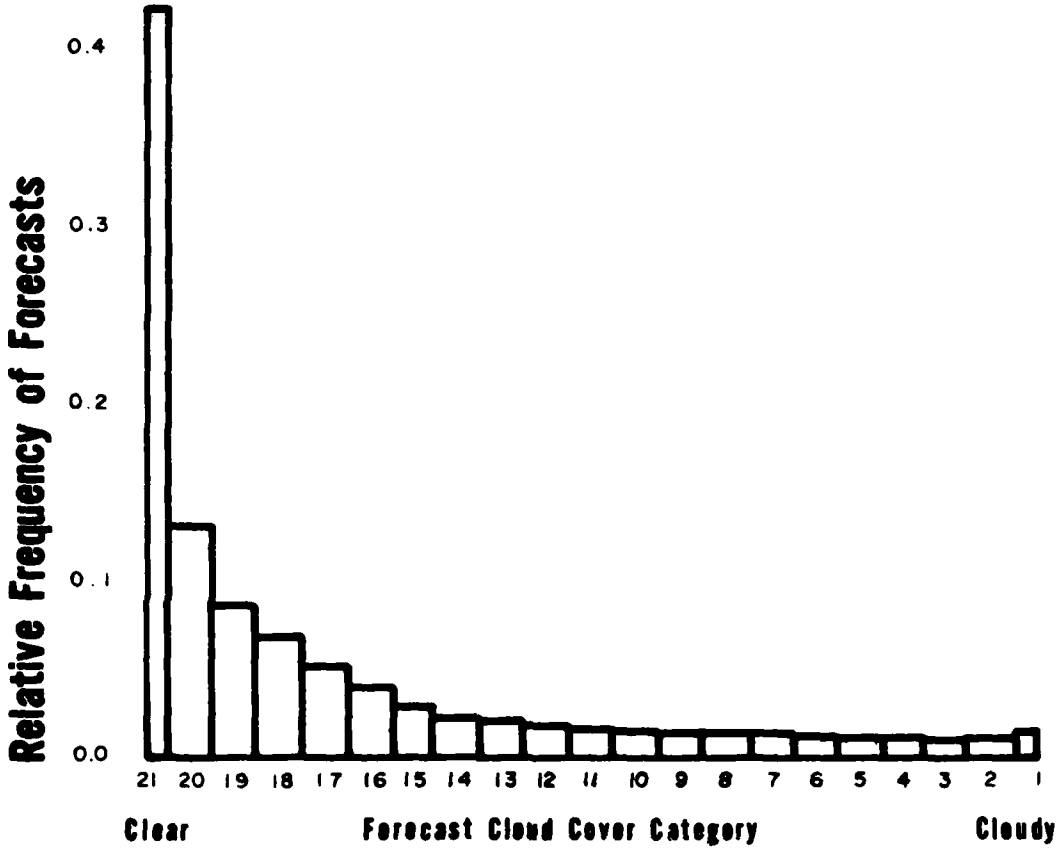


Figure 12. Skill Matrix Relative Frequencies of Forecasts Given a Forecast Month of January and an Observed (Verifying) Total Cloud Cover of Category 21 (Clear).

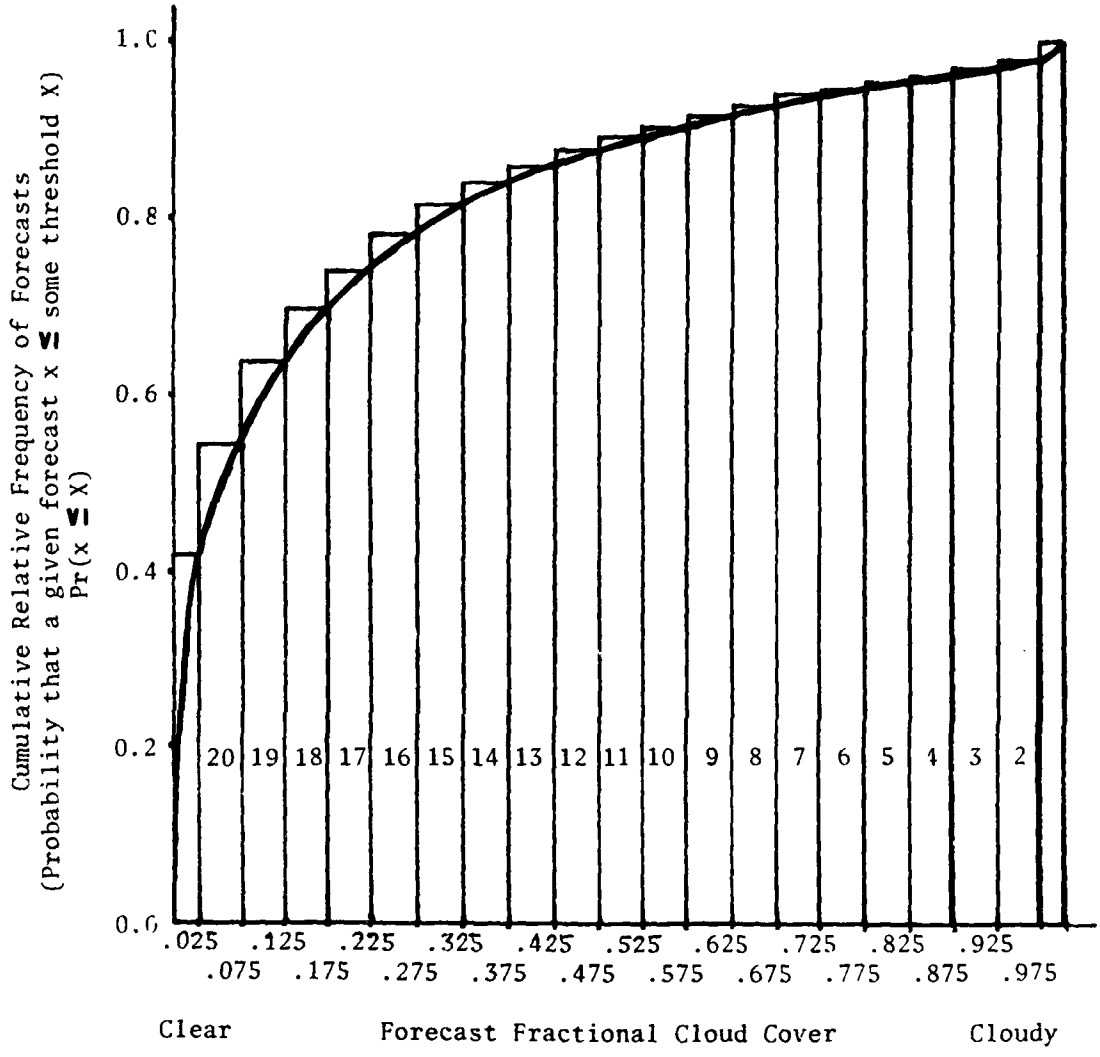


Figure 13. Skill Matrix Cumulative Relative Frequencies Given a Forecast Month of January and an Observed (Verifying) Total Cloud Cover of Category 21 (Clear).

Table 3. Skill Matrix Conditional Distribution of Forecasts for January, Observed Category 21 (Clear).

Forecast Category	Count	Relative Frequency	Cumulative Relative Frequency
21 (Clear)	20830	0.420	0.420
20	6506	0.131	0.551
19	4219	0.085	0.636
18	3254	0.066	0.702
17	2528	0.051	0.753
16	1883	0.038	0.791
15	1445	0.029	0.820
14	1106	0.022	0.842
13	940	0.019	0.861
12	769	0.016	0.877
11	684	0.014	0.891
10	623	0.012	0.903
9	602	0.012	0.915
8	550	0.011	0.926
7	464	0.010	0.936
6	501	0.010	0.946
5	463	0.009	0.955
4	424	0.009	0.964
3	469	0.009	0.973
2	587	0.012	0.985
1 (Cloudy)	747	0.015	1.000
TOTAL	49594		

where G and H are constants. Also, for a given observed category within the given month, the END of the observed cloud cover $\overset{''}{c}_o$ as computed by this model is a constant. This yields the conditional forecast simulation equation,

$$\overset{''}{c}_f = K + H\eta \tag{57}$$

where the constant K is equal to $G\,\overset{''}{c}_o$.

Given a known month and observed category, therefore, the conditional distribution of $\overset{''}{c}_f$ depends only on the distribution of η, which is normally distributed with a mean of zero and a variance of one, i.e., N(0,1). Equation (57) is used to produce random normal numbers $\overset{''}{c}_f$ with a mean of K and a variance of H^2 from a random normal number that is distributed N(0,1). Accordingly, the conditional forecast distribution $\overset{''}{c}_f$ produced by the simulation model is a normal distribution with a mean,

$$K = \rho_{fo} \tilde{c}_o \tag{58}$$

and variance,

$$H^2 = 1 - \rho_{fo}^2 \tag{59}$$

A sample conditional forecast distribution is shown in Figure 14.

Under these circumstances, the random normal numbers $\tilde{c}_f$ distributed $N(K,H^2)$ must be transformed to correspond to forecasts having an empirical distribution given by individual rows of the skill matrix, e.g., Table 3, Figures 12 and 13. This can be done by applying a compensation scheme that slightly adjusts the distribution of the output synthetic forecasts so as to reproduce the cumulative, row-normalized form of the skill matrix nearly exactly.

The basic concept of the adjustment scheme is to use the skill matrices themselves as inverse-normalizing transforms for the forecasts. The technique operates on the cumulative forms of both the skill matrix and model-generated conditional forecast distributions. Its logic is as follows.

Given a known month and observed category, each forecast category occupies a certain fixed interval (m,n) of the skill matrix cumulative relative frequency distribution. Since fractional cloud cover forecasts fall into this forecast category only if they are in the interval (i,j), where i and j are the fractional cloud cover boundaries for the forecast category, the skill matrix dictates that all conditional forecast cumulative probabilities which are in

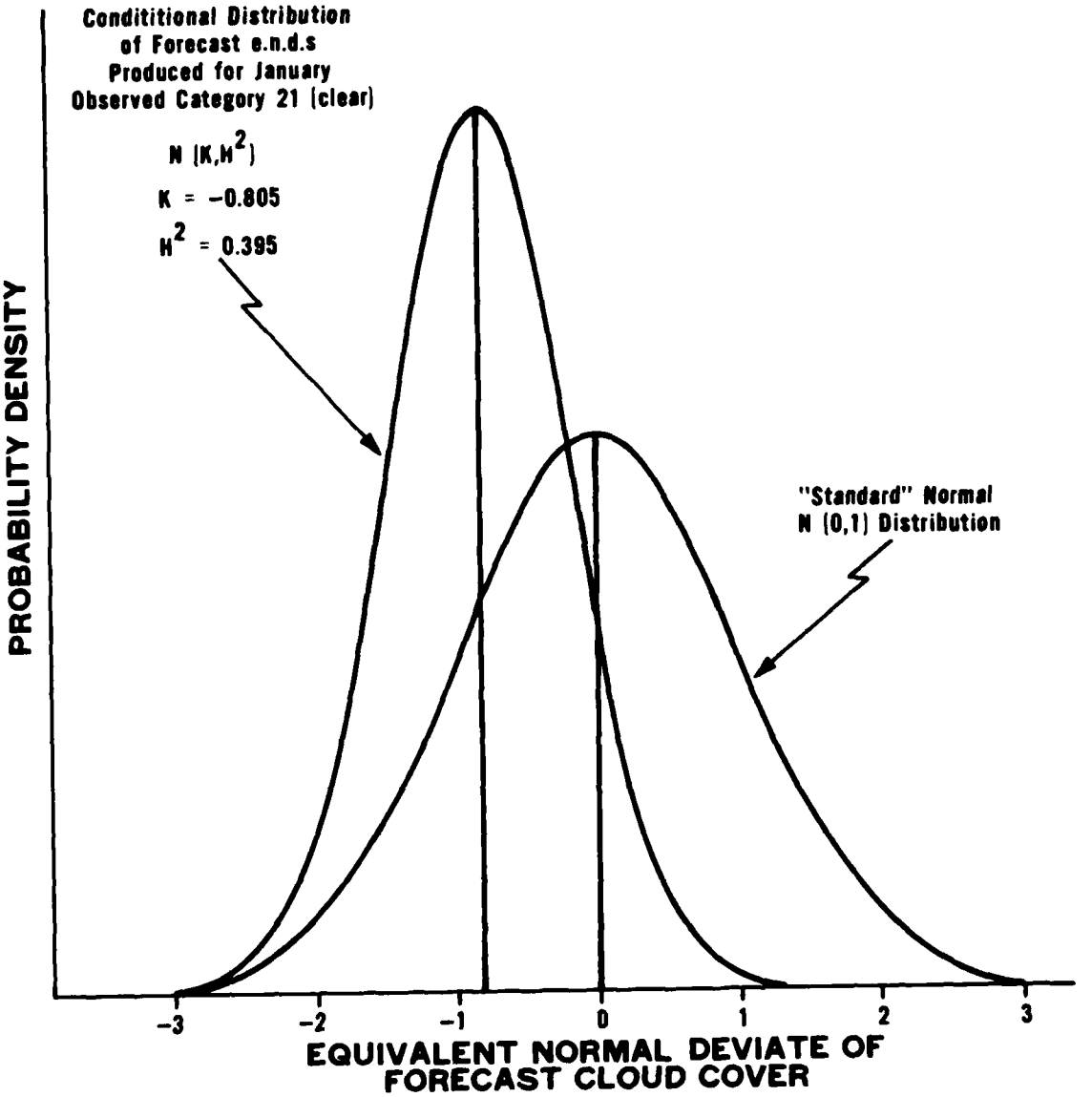


Figure 14. Sample Conditional Forecast Distribution Produced by the Cloud Forecast Simulation Equation Given a Known Month and Observed Category.

the interval (m,n) must be assigned forecast fractional cloud covers from the interval (i,j). These intervals are shown in Table 4 and are illustrated in Figure 15 for January, observed category 21 (clear).

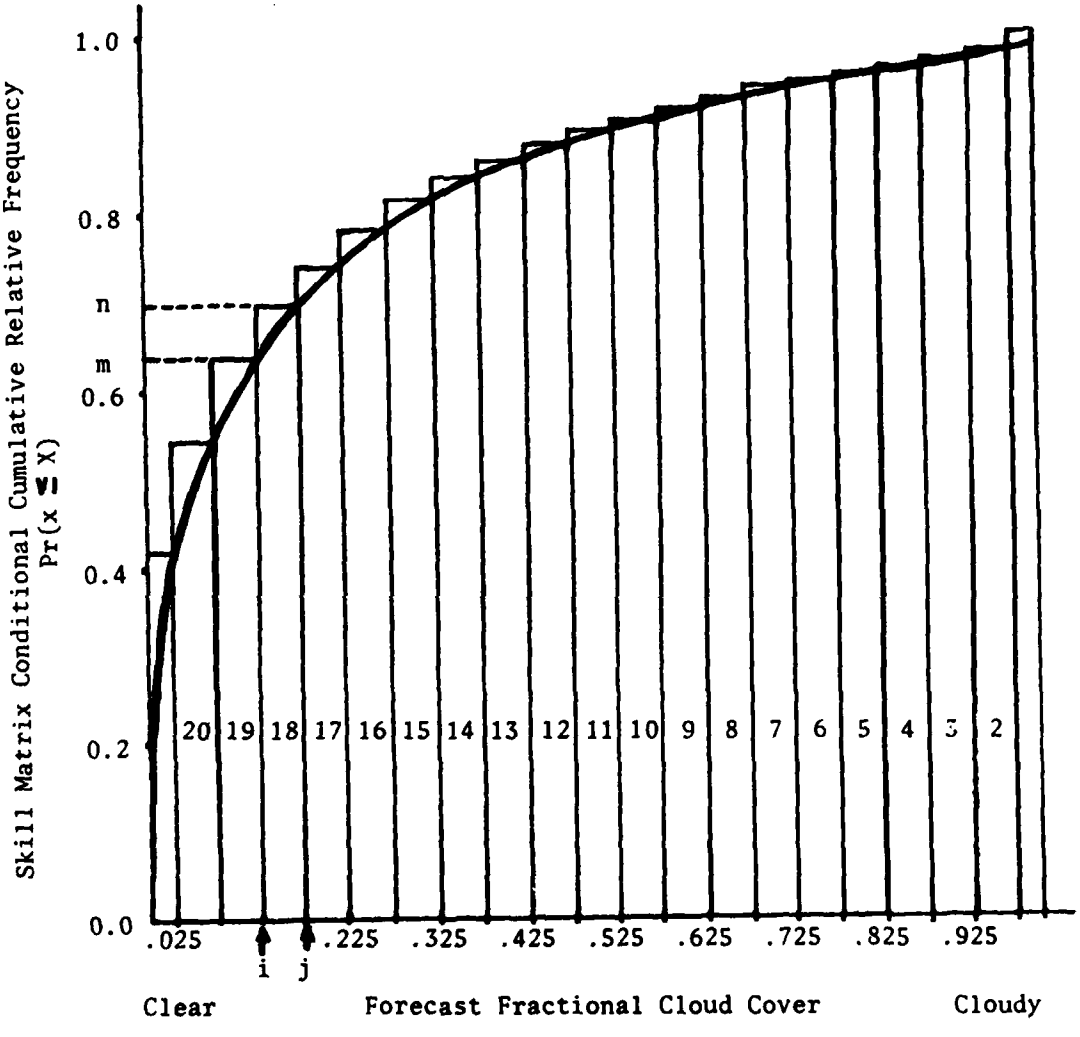


Figure 15. Mating of Skill Matrix Conditional Cumulative Relative Frequency Intervals (m,n) with Forecast Fractional Cloud Cover Intervals (i,j).

Table 4. Intervals Necessary to Insure Skill Matrix Reproduction Given Month January and Observed Category 21 (Clear).

Forecast Category	Character Representation in "Data Dump" Synthetic Forecast Field Displays	Forecast Fractional Cloud Cover Interval (i,j)	Skill Matrix Cumulative Relative Frequency Interval (m,n)	Intervals to Divide Pr( ) Distribution so as to Reproduce the Skill Matrix
21	0	0.000-0.025	0.000-0.420	0.000-0.420
20	1	0.025-0.075	0.420-0.551	0.420-0.551
19	2	0.075-0.125	0.551-0.636	0.551-0.636
18	3	0.125-0.175	0.636-0.702	0.636-0.702
17	4	0.175-0.225	0.702-0.753	0.702-0.753
16	5	0.225-0.275	0.753-0.791	0.753-0.791
15	6	0.275-0.325	0.791-0.820	0.791-0.820
14	7	0.325-0.375	0.820-0.842	0.820-0.842
13	8	0.375-0.425	0.842-0.861	0.842-0.861
12	9	0.425-0.475	0.861-0.877	0.861-0.877
11	A	0.475-0.525	0.877-0.891	0.877-0.891
10	B	0.525-0.575	0.891-0.903	0.891-0.903
9	C	0.575-0.625	0.903-0.915	0.903-0.915
8	D	0.625-0.675	0.915-0.926	0.915-0.926
7	E	0.675-0.725	0.926-0.936	0.926-0.936
6	F	0.725-0.775	0.936-0.946	0.936-0.946
5	G	0.775-0.825	0.946-0.955	0.946-0.955
4	H	0.825-0.875	0.955-0.964	0.955-0.964
3	I	0.875-0.925	0.964-0.973	0.964-0.973
2	J	0.925-0.975	0.973-0.985	0.973-0.985
1	K	0.975-1.000	0.985-1.000	0.985-1.000

To ensure that model-produced synthetic forecasts reproduce the skill matrix, therefore, requires that the conditional forecast distribution generated by the model be divided into 21 intervals such that the cumulative probabilities assigned to each interval, i.e., to each forecast category, are equal to the cumulative relative frequencies from the skill matrix for that month, observed category, and forecast category.

Now recall the model's conditional forecast simulation equation,

$$\underline{c}''_f = K + H\eta \tag{60}$$

Because K and H are constants, the conditional forecast distribution may be divided either directly (by dividing the $\underline{c}''_f$ distribution), or indirectly (by dividing the *random* normal η distribution). Dividing the η distribution eliminates the need to determine constants K and H for each month and observed category, thus, for simplicity, the η distribution is divided. The cumulative form of the η distribution, which is the cumulative normal probability distribution (denoted $Pr(\eta)$), is then easily divided into the intervals described above by using the skill matrix cumulative relative frequency intervals as cumulative normal probability intervals for each month and observed category. These intervals are shown in Table 4 and illustrated in Figure 16 for January, observed category 21 (clear).

With all intervals calculated and stored, the technique is applied to adjust a forecast field for a known month. As the technique operates on the cumulative form of the conditional forecast distribution, the spatially correlated random normal $\underline{\eta}$ field to be used by the scheme must first be

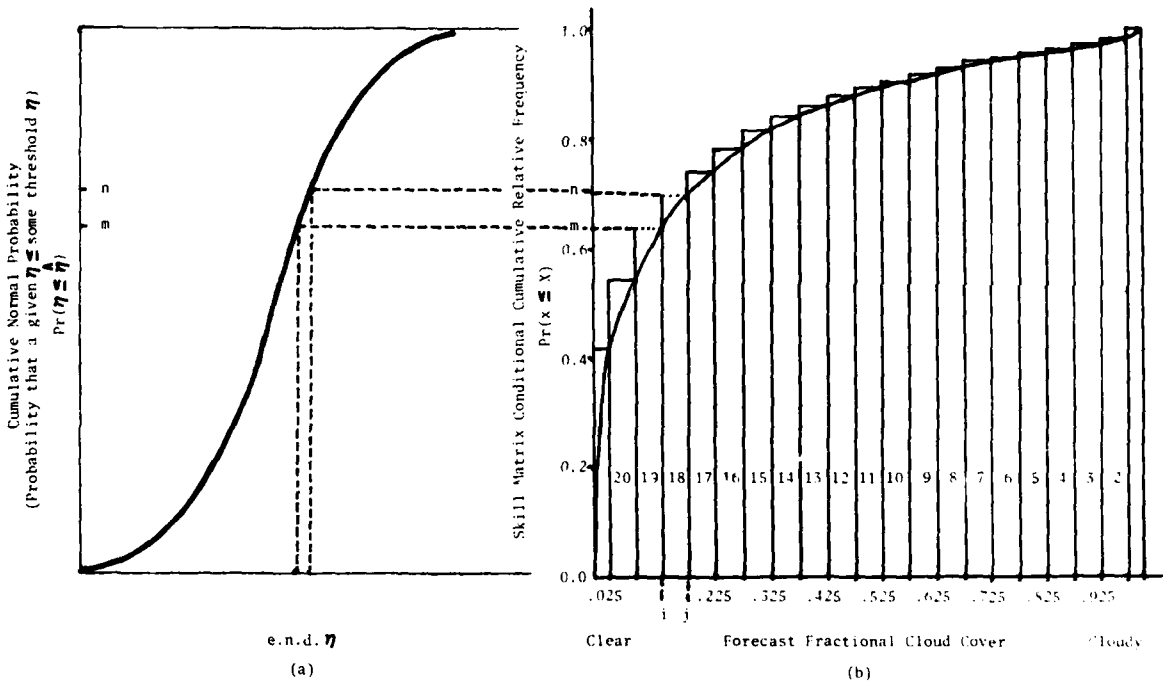


Figure 16. Linkage of Random Normal Distribution through the Cumulative Normal Probability Distribution and the Skill Matrix Conditional Cumulative Relative Frequency Distribution to the Forecast Fractional Cloud Cover Distribution.

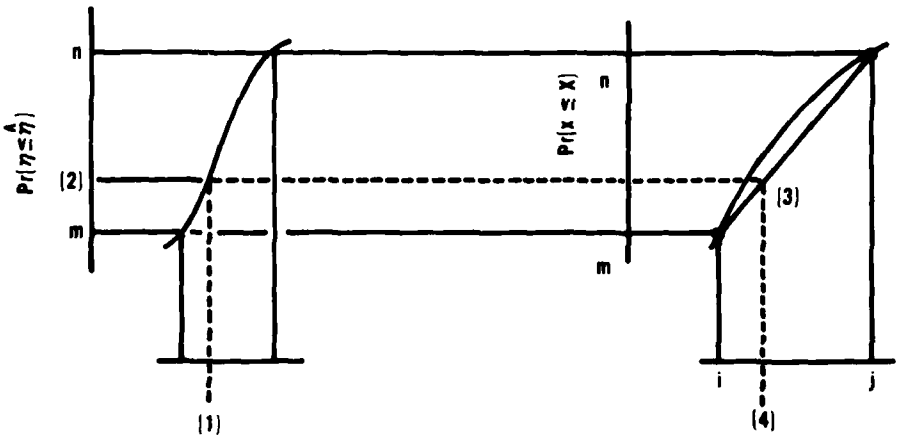
converted to its cumulative normal probability form Pr(η). At each grid point, the required integration of the normal distribution from -∞ to the END is easily accomplished by using a rational approximation.

The method then steps through each grid point, taking note of both the observed fractional cloud cover field c_o and the spatially correlated random normal END field in cumulative normal probability form Pr(η). At each grid point the observed cloud cover is evaluated to determine its observed category. Then the cumulative normal probability for this grid point from the Pr(η) field is tested against the conditional forecast distribution intervals previously

established to determine the forecast category at this grid point. If (as in operational practice) one needs to know only the forecast cloud cover category, then one needs to proceed no further. If a continuous forecast cloud cover value is needed (as in model development and testing), additional computations are required.

Figure 17 illustrates the method used to determine a continuous forecast cloud cover value. For the grid point under consideration, the technique first determines the fraction of the way the cumulative normal probability from the $\underline{Pr(\eta)}$ field is through its forecast category cumulative probability interval, which was determined above. This fraction is then used to interpolate linearly between the fractional cloud cover boundaries for the forecast category to a continuous fractional forecast cloud cover END for this grid point. This value is then stored in the forecast fractional cloud cover END field $\underline{c}_f$. Finally, the cloud cover forecast END for this grid point in the cloud cover forecast END field $\underline{c}''_f$ is post-adjusted to be that END which, if passed to the Johnson S_B inverse-normalizing technique, would have produced this grid point's forecast fractional cloud cover. These computations allow for statistical testing of the impact of the adjustment scheme on forecast distribution and spatial correlation.

The application of the adjustment scheme above is best illustrated by an example. Say the observed category for some grid point (I,J) in January is category 21 (clear). If the sawtooth wave model had produced a spatially correlated random END η of -0.0751 at grid point (I,J), the cloud forecast simulation model would have produced an unadjusted cloud cover forecast END c''_f of -1.0820 and an unadjusted forecast fractional cloud cover c_f of 0.0303. If the



(1) E.n.d. from spatially correlated $\underline{\eta}$ field.

(2) Cumulative normal probability of (1), from $\underline{Pr(\eta)}$ field.

–Determine the fraction of the way (2) is through (m,n). Formula

$$FRAC = \frac{(2) - m}{n - m}$$

(3) Use this fraction to linearly interpolate between fractional cloud cover boundaries (i,j).

(4) Compute continuous fractional cloud cover forecast c_f by the formula:

$$c_f = i + FRAC * (j - i)$$

Figure 17. Interpolation to a Continuous Cloud Cover Forecast.

adjustment scheme were applied, the $\eta(I,J)$ above would first be converted to its cumulative normal probability form $Pr(\eta)(I,J) = 0.4700$. This value lies between the cumulative normal probability boundaries (m,n) for forecast category 20. These boundaries are m = 0.4200 and n = 0.5512, as shown in Table 4. To arrive at the adjusted forecast fractional cloud cover, the technique first determines the fraction of the way the forecast is through its cumulative normal probability interval from the equation,

$$FRAC = \frac{Pr(\eta)(I,J) - m}{n - m} = \frac{0.4700 - 0.4200}{0.5512 - 0.4200} = 0.3811 \qquad (61)$$

The technique then linearly interpolates between fractional cloud cover boundaries (i,j) for forecast category 20 to an adjusted continuous forecast fractional cloud cover $c_f(I,J)$ by the equation,

$$c_f(I,J) = i + (FRAC)(j - i) \qquad (62)$$

$$c_f(I,J) = 0.025 + (0.3811)(0.075 - 0.025) = 0.0441$$

Finally, this forecast cloud cover is passed to the Johnson S_B inverse-normalizing technique to determine the adjusted $\overset{''}{c}_f(I,J,)$ which is found to be -0.9610.

3.7 Final Model: Hybrid Sawtooth Wave/Skill Matrix Design.

The final design of the Cloud Forecast Simulation Model, as displayed in Figure 18, consists of two submodels: a sawtooth wave submodel and an adjustment submodel. This version of the model accepts categorical* observed cloud cover fields and produces categorical synthetic forecast cloud cover fields which, after a sufficiently long run of the model, will precisely reproduce the set of input skill matrices.

*The development version of the Cloud Forecast Simulation Model produces continuous cloud cover values as well as cloud cover categories. This gave greater flexibility in development, but may not be needed operationally. In that case categorical inputs and outputs are sufficient.

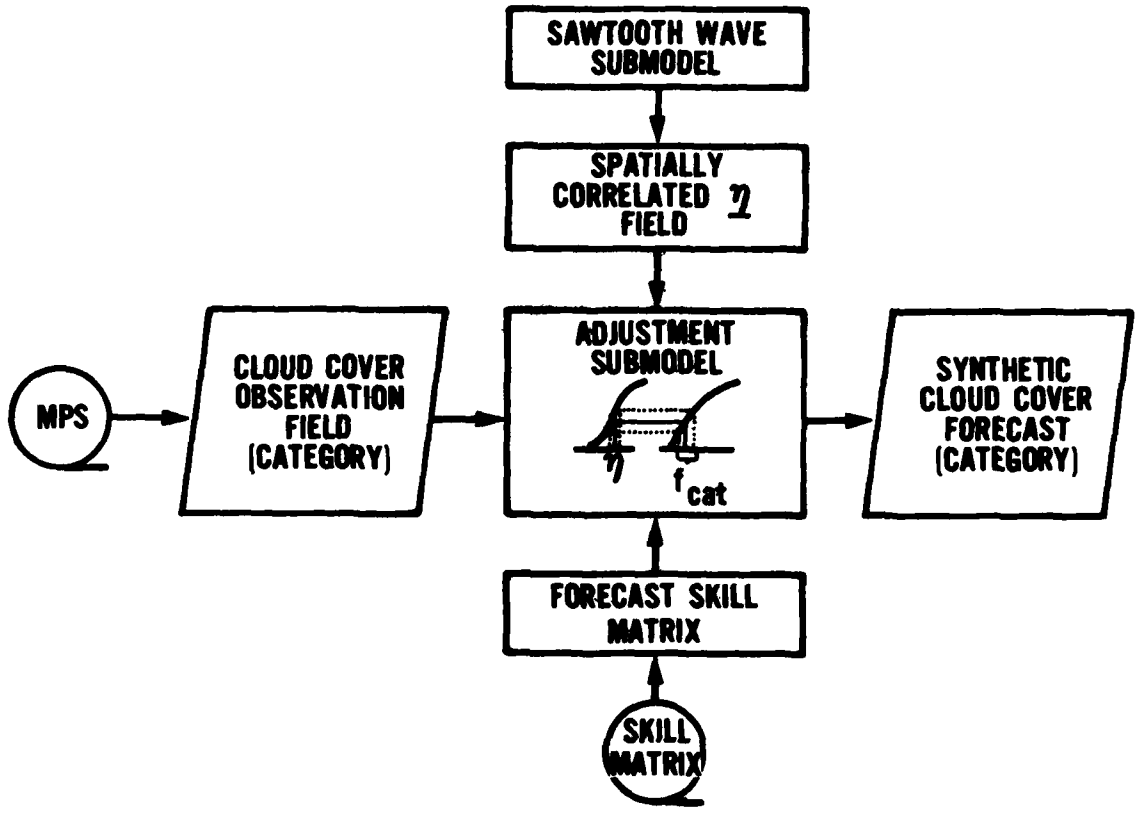


Figure 18. Final Model: Hybrid Sawtooth Wave/Skill Matrix Design.

The categorical observed cloud cover fields input to this version of the model are the same as those input to the earlier version, and are described in Section 3.4 of this report. The other input required by the model is a set of categorical skill matrices describing the performance of AFGWC's operational cloud forecasting model. These matrices can be input in either raw count or cumulative, row-normalized, percentage form, at the user's option.

When the model is run, the categorical skill matrices are first read in, converted to cumulative, row-normalized form, and stored. The model then begins reading observed cloud cover fields from an input file. The model produces as many synthetic forecast fields from each observed field as the user desires. For each of these synthetic forecast fields the sawtooth wave submodel generates one spatially correlated random normal number field $\underline{\eta}$. The adjustment submodel then operates on each grid point of the field, converting the η value to its cumulative normal probability form, comparing that probability against the stored skill matrix and determining the appropriate forecast cloud cover category for that grid point by use of the forecast adjustment scheme described in Section 3.6. After each grid point has been processed, the categorical synthetic forecast cloud cover field is displayed. This process is then repeated until the specified number of synthetic forecast fields for each observed field has been generated.

Chapter 4

SYNTHETIC SKILL OF THE CLOUD FORECAST SIMULATION MODEL

4.1 General.

One design goal of the Cloud Forecast Simulation Model is to produce synthetic forecasts of total cloud cover having about the same skill as cloud cover forecasts produced by an operational model. The operational product being modeled, in this case, is a set of 3-hour forecasts of total cloud cover, valid at 12 LST, drawn from Air Force Global Weather Central's (AFGWC's) SAVDOX cloud forecast data base for selected points.

AFGWC measures the skill of these cloud cover forecasts by accumulating 21 x 21-category forecast versus observation verification contingency tables for each month for both the Northern and Southern Hemispheres. The categories used range from 1 (cloudy) to 21 (clear), and are described in Table 5. The same categories are used for both forecasts and observations. The verification contingency tables, which will hereafter be referred to as "SAVDOX skill matrices," are built by verifying the category of each total cloud cover point forecast drawn from the SAVDOX cloud forecast data base against a categorical observed value interpolated from raw 3DNEPH data. One count is then added to the appropriate row (representing the observed category) and column (representing the forecast category) of the SAVDOX skill matrix for the month and hemisphere in which the forecast and observation are valid.

Table 5. Cloud Cover Categories for Verification of Total Cloud Cover Point Forecasts Drawn from AFGWC's SAVDOX Cloud Forecast Data Base.

Category	Cld-Cover-Intervals MPS	Simulation	Ascribed Midpoint Cld-Cover	Ascribed Fractional Equivalent Cld-Cover	Contrib Classes	Ascribed Fractional Upper Bound Cld-Cover
21	0%	0.0%-2.5%	1.25%	0.0125	21	0.025
20	1%- 5%	2.5%-7.5%	5.00%	0.0500	21-20	0.075
19	6%-10%	7.5%-12.5%	10.00%	0.1000	21-19	0.125
18	11%-15%	12.5%-17.5%	15.00%	0.1500	21-18	0.175
17	16%-20%	17.5%-22.5%	20.00%	0.2000	21-17	0.225
16	21%-25%	22.5%-27.5%	25.00%	0.2500	21-16	0.275
15	26%-30%	27.5%-32.5%	30.00%	0.3000	21-15	0.325
14	31%-35%	32.5%-37.5%	35.00%	0.3500	21-14	0.375
13	26%-40%	37.5%-42.5%	40.00%	0.4000	21-13	0.425
12	41%-45%	42.5%-47.5%	45.00%	0.4500	21-12	0.475
11	46%-50%	47.5%-52.5%	50.00%	0.5000	21-11	0.525
10	51%-55%	52.5%-57.5%	55.00%	0.5500	21-10	0.575
9	56%-60%	57.5%-62.5%	60.00%	0.6000	21-9	0.625
8	61%-65%	62.5%-67.5%	65.00%	0.6500	21-8	0.675
7	66%-70%	67.5%-72.5%	70.00%	0.7000	21-7	0.725
6	71%-75%	72.5%-77.5%	75.00%	0.7500	21-6	0.775
5	76%-80%	77.5%-82.5%	80.00%	0.8000	21-5	0.825
4	81%-85%	82.5%-87.5%	85.00%	0.8500	21-4	0.875
3	86%-90%	87.5%-92.5%	90.00%	0.9000	21-3	0.925
2	91%-95%	92.5%-97.5%	95.00%	0.9500	21-2	0.975
1	96%100%	97.5%-100%	98.75%	0.9875	All	1.000

Two points regarding this process are worth noting. First, forecasts drawn from the SAVDOX cloud forecast data base are made by smoothing the resident 1/8-mesh forecast data to 1/4-mesh before interpolating to the selected forecast point. On the other hand, the observations used to verify these forecasts are interpolated from raw (unsmoothed), 1/8-mesh 3DNEPH data. Smoothing has

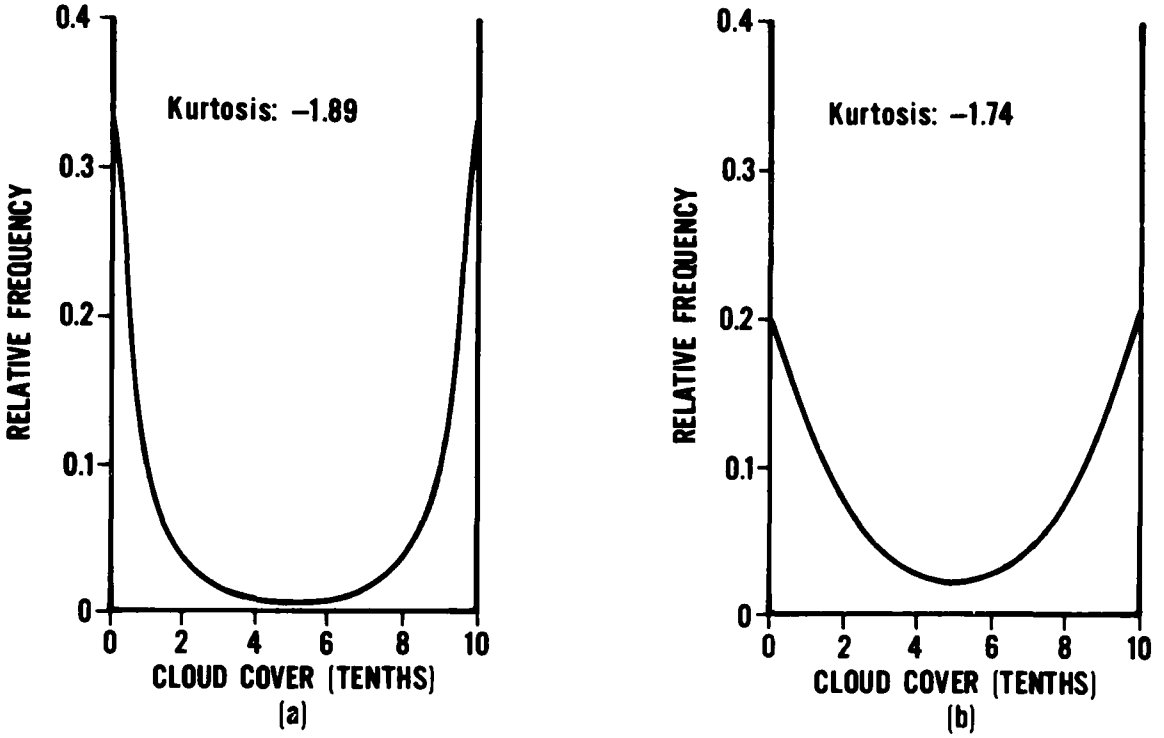


Figure 19. Kurtosis in Cloud Cover Distributions. Increasing the relative frequency of clear (0 tenths) and overcast (10 tenths), which are already the modes of the bimodal, U-shaped cloud cover distribution, has the effect of decreasing the kurtosis of the distribution, i.e., making it more negative (more platykurtic). Case (a) might be the distribution of forecast cloud cover before smoothing, and (b) might be the distribution after smoothing. Characteristically, smoothing makes cloud cover distributions less platykurtic.

the effect of decreasing the relative frequency of clear and overcast and filling in the middle of the probability distribution. As shown in Figure 19, this effect of smoothing can be described in terms of making the smoothed distribution less platykurtic than the unsmoothed distribution. Because the skill matrices are formed from smoothed forecasts and unsmoothed verifying observations, the marginal probability distributions of observed cloud cover from the skill matrices will be more platykurtic than the marginal distributions of

forecasts from the same matrices. This is the case in the SAVDOX skill matrices for all months for both hemispheres. An example for the Northern Hemisphere in January is given in Figure 20. Because the Cloud Forecast Simulation Model is designed to simulate the operational product, it should, and does, produce forecast marginal distributions which are less platykurtic than the distributions of input observations.

Second, after the counts of cases in the SAVDOX skill matrices are accumulated, the matrices are referred to as being in raw count form (see Table 2 for an example). These raw count matrices are then analyzed to determine the marginal probability distributions of forecasts and observations. The elements of the marginal distribution of observations for each month and hemisphere are then used to normalize each row of their respective raw count skill matrices, so that the sum of all matrix elements across each row is one. The matrix elements for each row are then accumulated from clear to cloudy to yield the cumulative, row normalized form of each SAVDOX skill matrix (see Table 6 for an example). When the Cloud Forecast Simulation Model is used in practice, it is fed a field of Multi-purpose Simulator (MPS) data as input observations. These input MPS data are not likely to be distributed exactly as were the observed data used to make up the skill matrices. Because of this difference between the marginal probability distribution of input observations and that inherent in the skill matrix, the model cannot be expected to reproduce the skill matrices in raw count form. Rather, the model is designed to reproduce the conditional forecast distributions for each month, hemisphere, and observed category, as defined by the rows of the cumulative, row-normalized SAVDOX skill matrices.

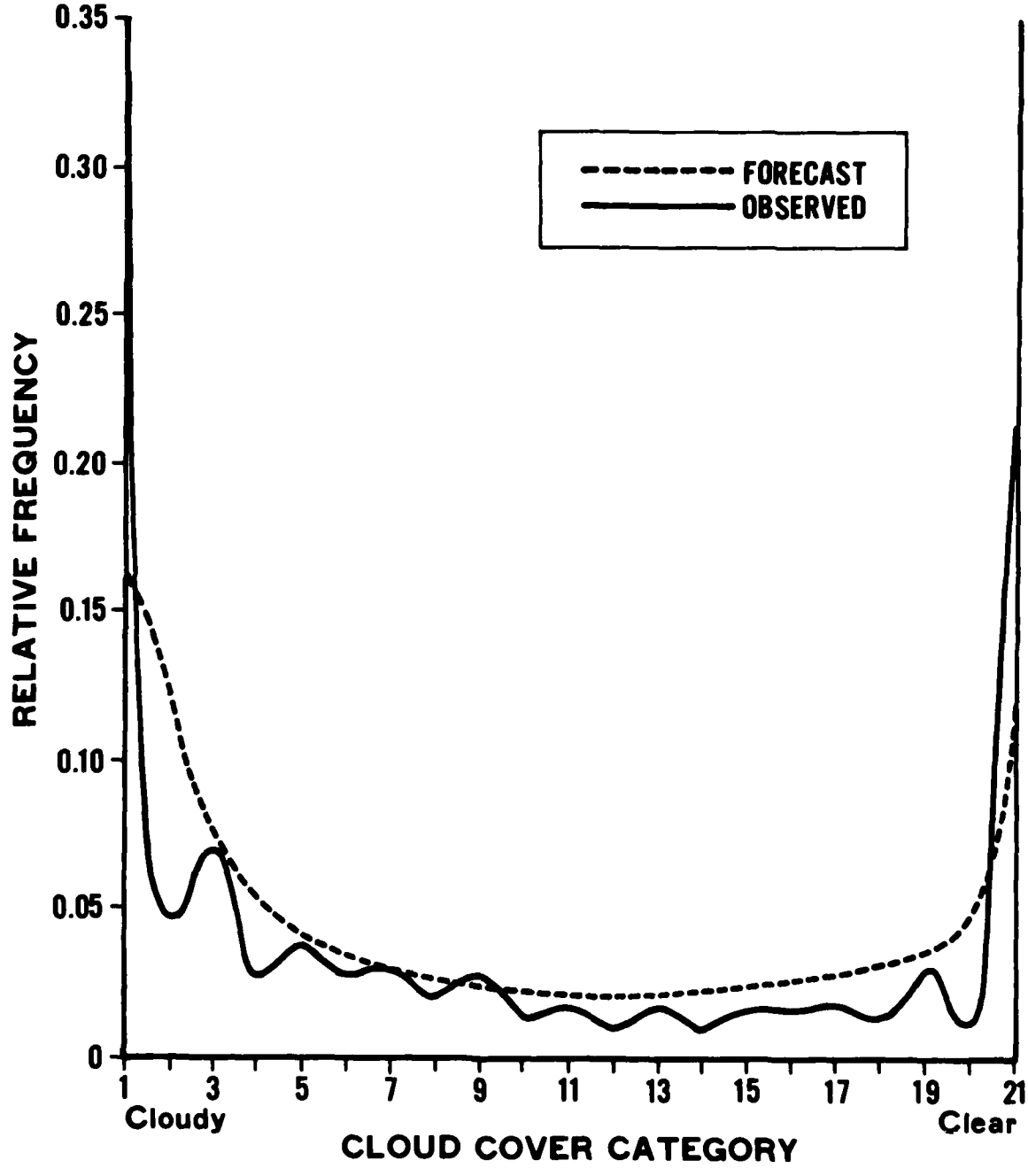


Figure 20. Marginal Distributions of Forecast and Observed Total Cloud Cover from the AFGWC SAVDOX Skill Matrix for January, Northern Hemisphere.

Table 6. Cumulative, Row-normalized, Percentage AFGWC SAVDOX Skill Matrix for January Northern Hemisphere. The cumulative, row-normalized skill matrix elements are multiplied by 100 and rounded to integer to create this form of the SAVDOX skill matrix.

		CLOUDY								FORECAST											CLEAR
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	1	100	65	42	30	23	19	15	13	11	10	8	7	6	5	4	4	3	2	2	1	1
	2	100	80	60	47	38	32	27	23	20	18	15	13	11	10	8	7	6	5	3	2	1
	3	100	84	67	55	46	39	34	30	26	23	20	18	15	13	11	9	8	6	5	3	2
	4	100	88	74	63	54	47	40	35	31	28	25	22	19	17	14	12	10	8	6	4	2
	5	100	88	75	65	56	50	44	40	35	32	28	25	22	19	17	14	12	9	7	5	3
	6	100	90	78	69	62	54	49	44	40	36	32	28	25	22	19	17	14	11	9	6	4
O	7	100	90	78	70	62	56	50	45	41	37	33	30	26	23	20	17	14	12	9	6	4
B	8	100	89	78	69	63	57	52	47	42	38	34	30	27	24	20	18	15	12	9	6	4
S	9	100	91	81	73	66	60	55	50	46	42	38	34	31	27	24	21	17	14	11	8	5
E	10	100	92	84	78	73	67	63	58	54	49	46	42	37	34	29	25	22	18	14	10	6
R	11	100	96	90	85	80	75	71	66	61	57	52	47	43	38	33	29	24	20	15	12	7
V	12	100	95	90	86	82	77	73	69	65	61	56	51	47	42	37	32	28	23	18	14	9
E	13	100	96	92	88	85	80	77	73	69	65	61	57	52	47	42	37	31	26	21	16	9
D	14	100	97	93	89	87	83	80	76	72	67	64	60	55	50	44	40	34	29	23	17	11
	15	100	97	93	89	86	83	81	78	74	71	67	63	58	54	48	42	36	31	25	19	11
	16	100	97	94	91	87	84	81	79	76	73	70	66	62	57	53	48	43	37	30	22	14
	17	100	97	95	92	90	87	85	82	79	77	74	71	66	62	58	52	46	39	32	24	15
	18	100	98	95	93	92	90	87	85	83	80	78	75	71	68	63	58	51	44	36	27	18
	19	100	98	96	95	93	91	89	87	86	83	81	78	76	73	69	64	59	52	44	35	24
	20	100	97	95	93	91	89	87	86	83	81	79	76	74	71	68	63	58	51	44	34	22
	21	100	98	97	96	96	95	94	93	92	90	89	88	86	84	82	79	75	70	64	55	42

Total Observations: 232893
Total Forecasts: 232893
Tetrachoric Slice Pts (I , J): 4 6
Tetrachoric Rgns: A , B: 88304 23004
C , D: 27502 94083
Tetrachoric Fcst-Obs Corr: 0.777575

The fact that these conditional forecast distributions are reproduced at each grid point, independent of the cloud cover climatology of that grid point, is a major limiting factor of the present Cloud Forecast Simulation Model. Because the SAVDOX skill matrices are built by verifying forecasts for points irregularly distributed over an entire hemisphere, the conditional forecast distributions defined by each row of these skill matrices describe only the composite performance of the forecasts for all forecast points, and are not representative of the forecast performance at each individual point in the set. Since the actual conditional forecast distributions are likely to be quite different for forecast points in the climatically different regions that went into building the hemispheric composite SAVDOX skill matrices (e.g., desert, tropics, etc.), the forecasts for an individual grid point or region may be artificially clear (the tropics) or cloudy (the desert), as long as only one matrix is used for an entire hemisphere. Given these known limitations, testing was begun to determine whether the Cloud Forecast Simulation Model could reproduce the skill of AFGWC's operational forecasts, as described by the SAVDOX skill matrices in cumulative, row-normalized form.

4.2 Skill Testing of the Early Model.

Initial testing of the Cloud Forecast Simulation Model was done on a 10 x 10-point, AFGWC 1/4-mesh (50-NM resolution) development subgrid using the observations for January 1976. This small subgrid was chosen so as not to require extensive data processing resources for the initial testing of the model.

In the first test run, the input parameters, the tetrachoric forecast-observation correlation (ρ_{fo}) and the Johnson S_B normalizing coefficients for observed (a,b) and forecast (c,d) cloud cover distributions, were those computed from the raw count SAVDOX skill matrix for January for the Northern Hemisphere, and the analysis of its marginal distributions. Because the observed cloud cover distribution for the development subgrid for January 1976 was much less clear than that which went into building the January, Northern Hemisphere SAVDOX skill matrix, the normalizations of observed and forecast cloud cover were biased. This resulted in mean equivalent normal deviates (ENDs) of observed and forecast cloud cover being different from the theoretical value of zero. It also resulted in the standard deviations of the ENDs of observed and forecast cloud cover being much smaller than the theoretical value of 1. The tetrachoric forecast-observation correlation of the model-produced synthetic skill matrix was far less than the input ρ_{fo}. And the synthetic, cumulative, row-normalized, percentage skill matrix produced by a long run of the model was different from the SAVDOX matrix in the same form (see Table 7).

In the next step of model testing, an attempt was made to correct for the differences between the observed cloud cover distribution for the development subgrid for January 1976 and the observed distribution from the January, Northern Hemisphere SAVDOX skill matrix. The Johnson S_B normalizing coefficients (a and b) for the observed cloud cover distribution, which had been computed from the SAVDOX matrix in the previous test, were replaced with coefficients computed expressly for the observed cloud cover distribution of the development subgrid for January 1976. This test produced a synthetic ρ_{fo} much closer to the input ρ_{fo}, and observed and forecast mean ENDs closer to zero and

Table 7. Synthetic Skill Matrix Produced by the First Run of the Cloud Forecast Simulation Model for January 1976. The model's input parameters were those computed from the January Northern Hemisphere AFGWC SAVDOX skill matrix.

		CLOUDY							FORECAST												CLEAR
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	1	100	74	58	50	45	40	36	33	30	28	25	23	21	19	16	14	12	10	8	5	2
	2	100	80	66	58	53	48	44	41	38	35	32	30	27	25	22	20	17	14	11	8	3
	3	100	83	70	63	57	52	49	45	42	39	36	33	31	28	25	23	20	17	13	9	4
	4	100	85	72	65	60	55	51	48	45	42	39	36	33	30	28	25	22	19	15	11	5
	5	100	86	74	67	62	57	53	50	47	44	41	38	35	32	29	26	23	20	16	12	5
	6	100	87	75	69	63	59	55	51	48	45	42	39	37	34	31	28	25	21	17	12	6
O	7	100	87	77	70	65	60	56	53	50	47	44	41	38	35	32	29	26	22	18	13	6
B	8	100	88	78	71	66	61	58	54	51	48	45	42	39	36	33	30	27	23	19	14	7
S	9	100	89	79	72	67	63	59	55	52	49	46	43	40	37	34	31	28	24	20	15	7
E	10	100	89	79	73	68	64	60	57	53	50	47	44	42	39	36	32	29	25	21	15	7
R	11	100	90	80	74	69	65	61	58	55	51	49	46	43	40	37	33	30	26	22	16	8
V	12	100	90	81	75	70	66	62	59	56	53	50	47	44	41	38	34	31	27	22	17	8
E	13	100	91	82	76	71	67	63	60	57	54	51	48	45	42	39	36	32	28	23	17	9
D	14	100	91	83	77	72	68	65	61	58	55	52	49	46	43	40	37	33	29	24	18	9
	15	100	92	83	78	73	69	66	62	59	56	53	50	48	45	41	38	34	30	25	19	10
	16	100	92	84	79	74	71	67	64	61	58	55	52	49	46	43	39	36	31	26	20	10
	17	100	93	85	80	76	72	69	65	62	59	57	54	51	48	44	41	37	33	28	21	11
	18	100	94	86	81	77	74	70	67	64	61	58	56	53	50	46	43	39	35	30	23	12
	19	100	94	88	83	79	76	73	70	67	64	61	58	55	52	49	46	42	37	32	25	14
	20	100	96	90	86	82	79	76	73	71	68	65	62	60	57	53	50	46	41	36	28	16
	21	100	97	93	90	87	84	82	80	77	75	72	70	67	65	61	53	54	50	44	36	22

Total Observations: 310000
Total Forecasts: 310000
Mean Obs E.N.D.: 0.06116
Mean Fcst E.N.D.: 0.04755
Std Dev Obs E.N.D.: 0.40559
Std Dev Fcst E.N.D.: 0.70306
Tetrachoric Slice Pts (I,J): 3 6
Tetrachoric Rgns A, B : 91163 61046
C, D : 62866 94925
Tetrachoric Fcst-Obs Corr: 0.30978
Pearson Product-Moment Corr of E.N.D.s: 0.44851

standard deviations closer to one. However, the synthetic, cumulative, row-normalized skill matrix produced by a long run of the model looked even less like the corresponding SAVDOX matrix than the synthetic matrix of the previous test. With this result, the small subgrid (10 x 10-point development) was abandoned in favor of an enlarged grid encompassing an area having an observed distribution similar to that which went into building the SAVDOX skill matrices.

The development subgrid was accordingly enlarged to a 97 x 80-point, 1/4-mesh grid that covered most of the Eurasian continent. The extent of this enlarged development subgrid is shown in Figures 9-11. The first test run using this enlarged subgrid was made using cloud cover data for January 1979 and input parameters as follows: ρ_{fo} from the January, Northern Hemisphere SAVDOX skill matrix; Johnson S_B coefficients a and b from the January 1979 observed cloud cover distribution for the 97 x 80 subgrid; and c and d from a subjectively derived, hypothetical forecast distribution that attempted to account for the differences in the two observed distributions. The results of this test were very encouraging. Because the observed distribution for January 1979 for the enlarged subgrid was much more like that which went into building the January, Northern Hemisphere SAVDOX skill matrix, the only major discrepancy that appeared in the model output was a curious bias against clear forecasts in each synthetic conditional forecast distribution (see Table 8). The encouragement was short-lived, however, as efforts to reduce this bias by adjusting the model's input parameters proved futile.

This effort to reduce the model's bias did, however, lead to the eventual discovery of the reasons for that bias. In this effort, the concept of using

Table 8. Synthetic Skill Matrix Produced by the Early Cloud Forecast Simulation Model for January 1979, for the 97 x 80 Subgrid. The Johnson SB normalizing coefficients input to the model were calculated specifically for the observed and "consistent" forecast distributions of the subgrid for this mont...

		CLOUDY	FORECAST																		CLEAR
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	1	100	66	43	32	25	20	16	13	11	9	8	6	5	4	3	2	1	2	1	0	0
	2	100	81	61	49	41	35	30	26	22	19	17	14	12	10	8	6	5	3	2	1	0
	3	100	87	70	59	51	44	39	34	30	27	23	20	18	15	12	10	8	6	4	2	0
	4	100	90	75	65	57	50	45	40	36	32	28	25	22	19	16	13	10	8	5	3	1
	5	100	91	78	69	61	55	49	44	40	36	32	29	25	22	19	16	13	9	6	4	1
	6	100	93	81	72	65	58	53	48	44	40	36	32	28	25	21	18	15	11	8	4	1
O	7	100	94	83	75	68	62	56	51	47	43	39	35	31	27	24	20	16	13	9	5	1
B	8	100	95	85	77	70	64	59	54	50	46	42	38	34	30	26	22	18	14	10	6	2
S	9	100	95	86	79	73	67	62	57	53	48	44	40	36	32	28	24	20	16	12	7	2
E	10	100	96	88	81	75	69	64	60	55	51	47	43	39	35	31	26	22	18	13	8	2
R	11	100	97	89	82	77	72	67	62	58	54	49	45	41	37	33	29	24	19	14	9	3
V	12	100	97	90	84	79	74	69	65	60	56	52	48	44	40	35	31	26	21	16	10	3
E	13	100	97	91	86	81	76	71	67	63	59	55	51	46	42	38	33	28	23	18	11	3
D	14	100	98	92	87	82	78	74	70	65	61	57	53	49	45	40	36	31	25	19	12	4
	15	100	98	93	89	84	80	76	72	68	64	60	56	52	48	43	39	33	28	21	14	5
	16	100	99	94	90	86	82	78	75	71	67	63	59	55	51	47	42	36	31	24	16	6
	17	100	99	95	92	88	85	81	78	74	70	67	63	59	55	50	45	40	34	27	18	7
	18	100	99	96	93	90	87	84	81	78	74	71	67	63	59	55	50	44	38	31	21	8
	19	100	99	97	95	93	90	87	85	82	79	76	72	69	65	61	56	50	44	36	26	11
	20	100	100	99	97	95	94	92	90	87	85	83	80	77	73	70	65	60	54	45	34	16
	21	100	100	100	99	98	98	97	96	95	93	92	90	88	86	83	80	76	71	64	52	30

Total Observations: 2405600
Total Forecasts: 2405600
Tetrachoric Slice Pts (I,J): 11 10
Tetrachoric Rgns A, B : 918927 278103
C, D : 255514 953056
Tetrachoric Fcst-Obs Corr: 0.76702

Mean Obs E.N.D.: 0.00430
Mean Fcst E.N.D.: 0.00332

Std Dev Obs E.N.D.: 0.76043
Std Dev Fcst E.N.D.: 0.86283

Pearson Product-Moment Corr of E.N.D.s: 0.68534

fixed input parameters, even for a hemispheric grid, was abandoned, and a technique was developed (program FCLDJ) to compute the optimum input parameters for any observed cloud cover distribution. The technique assumes that the cumulative, row-normalized form of each SAVDOX skill matrix will be reproduced exactly by the model. Given a known observed distribution, therefore, the interior of the skill matrix and one margin are known. The remaining margin, the "consistent" forecast distribution, is then uniquely defined and can be easily computed. The observed and consistent forecast distributions are then fitted with Johnson S_B normalizing curves, and optimum Johnson coefficients a, b, c, and d are determined. The remaining input parameter, ρ_{fo}, is computed by using the observed distribution and the appropriate cumulative, row-normalized SAVDOX skill matrix to construct an idealized raw count matrix for the input distribution. This consistent raw count matrix is then analyzed to determine its tetrachoric forecast-observation correlation, ρ_{fo}, and the set of optimum input parameters is complete.

A by-product of this input parameter optimizing technique was the discovery that the synthetic skill matrix produced by the Cloud Forecast Simulation Model after a long run could be solved for analytically. Expected synthetic skill matrices computed by this procedure also showed the bias against clear forecasts and helped point out the two fundamental reasons for it.

The first reason is a deficiency in the Johnson S_B normalizing transforms of observed and forecast cloud cover values. These normalizations consistently produce fitted forecast and observed cloud cover distributions with only half to three quarters of the clear cloud cover cases in the actual distributions (see Table 9).

Table 9. Percentage of Clear Cloud Cover Cases from the Northern Hemisphere AFGWC SAVDOX Skill Matrix Observed and Forecast Cloud Cover Distributions Which Are Found in their Corresponding Johnson S_B Fitted Distributions.

	PERCENTAGES	
Month	Forecast	Observed
Jan	77	75
Feb	75	76
Mar	71	71
Apr	71	72
May	68	68
Jun	53	66
Jul	51	53
Aug	62	64
Sep	72	68
Oct	75	71
Nov	77	71
Dec	80	73

Example showing computation of value 77 percent for January, Northern Hemisphere forecast distribution. All variables used in this computation were calculated from the January, Northern Hemisphere SAVDOX skill matrix.

Total cases in skill matrix	= 232893
Total cases with observed category of 21 (clear)	= 28611
Relative frequency of clear cases in skill matrix = 28611/232893	= 0.1229
Relative frequency of clear cases in Johnson SB fitted distribution	= 0.0951
Percentage of clear cases from the skill matrix found in the Johnson S_B fitted distribution = 0.0951/0.1229	= 0.77

This fitting problem undoubtedly had a large influence in producing the bias against clear forecasts noted above, but was not an insurmountable problem in itself. The second reason, however, proved to be the death knell for the early version of the Cloud Forecast Simulation Model. This second reason is that Equation (4), the heart of the early model, assumes a bivariate normal distribution of forecasts and observations. Experiments showed, however, that even with perfect normalizations of the forecasts, the observed distributions could not be normalized in a manner which would reproduce the SAVDOX skill matrices; i.e., the SAVDOX skill matrices are not bivariate normal. This is not unreasonable because the skill matrices represent a composite of many

potentially differing climatologies. The joint distribution of forecasts and observations from the skill matrices is thus a weighted mix of bivariate normal distributions and is not in itself bivariate normal. Because assuming the matrices are bivariate normal did not produce synthetic skill matrices close enough to the SAVDOX matrices to meet customer needs, the early, or unadjusted, version of the model was abandoned, and a new approach was taken.

4.3 Skill Testing of the Final Model.

This new approach resulted in development of the forecast adjustment scheme described in Section 3.6 of this report. With the application of this scheme, both of the problems in the early, unadjusted version of the model were circumvented. First, since in the adjustment scheme the cumulative, row-normalized forms of the SAVDOX skill matrices are themselves used as normalizing transforms, the Johnson S_B normalizing transforms are not needed at all. Second, the bivariate normal assumption is unnecessary because, instead of using Equation (4), which operates on each skill matrix as a whole, the adjustment scheme operates on the individual rows of each skill matrix. Because each row describes forecast performance contingent upon the observed category at the particular forecast point, the only assumption necessary is that the model-produced $\eta(i,j)$ fields are distributed random normally, a condition which depends only on the goodness of the computer's uniform random number generator and the approximation that the sum of N uniform random numbers approaches a normally distributed random number.

Tests of the adjusted model for the 97 x 80 subgrid for January 1979 and the midseason months of 1977 (January, April, July, and October) showed good results. In each case the appropriate cumulative, row-normalized, percentage SAVDOX skill matrix was reproduced to within 1 percent at each element by a long run of the model. An example of the synthetic, cumulative, row-normalized, percentage skill matrix produced by the final form of the model for January 1979 is shown in Table 10.

4.4 Limitations and Recommendations.

Despite the model's success in reproducing SAVDOX skill matrices, it should be kept in mind that those matrices themselves represent one of the most important limitations of the model. With the present model design, the conditional forecast distributions from the AFGWC SAVDOX skill matrices will be reproduced at each grid point in the hemisphere, regardless of the cloud cover climatology of that grid point. There are two possible ways to minimize this limitation.

The first would be to generate SAVDOX skill matrices for regions smaller than an entire hemisphere. If these matrices were generated for regions of essentially homogeneous cloud cover climatology and were sufficiently well populated to adequately define the performance of the operational cloud forecast model, the composite conditional forecast distributions from the skill matrices would be much closer to the actual conditional forecast distributions for each grid point in the region. If neighboring regions have significantly differing cloud cover climatologies, however, this method may create artificial discontinuities in the forecast cloud cover fields along regional boundaries.

Table 10. Cumulative, Row-normalized, Percentage Synthetic Skill Matrix Produced by a Long Run of the Final Cloud Forecast Simulation Model for January 1979, for the 97 x 80 Subgrid. Each element is within 1 percent of its corresponding element from the AFGWC SAVDOX skill matrix in this form (Table 6).

		CLOUDY	FORECAST																		CLEAR
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	1	100	65	42	31	24	19	16	13	12	10	9	7	6	5	4	4	3	2	2	1	1
	2	100	80	60	47	39	32	27	24	21	18	16	14	12	10	8	7	6	5	4	3	1
	3	100	84	67	56	46	40	34	30	26	23	20	18	16	13	11	10	8	6	5	3	2
	4	100	88	74	63	54	47	41	36	32	28	25	22	19	17	14	12	10	8	6	4	2
	5	100	88	75	65	56	50	44	40	36	32	29	25	23	20	17	14	12	9	7	5	3
	6	100	90	78	69	62	54	49	44	40	37	32	29	26	23	20	17	14	11	9	6	4
O	7	100	90	78	70	63	56	50	45	41	37	33	30	27	23	20	17	14	12	9	6	4
B	8	100	89	78	69	63	57	52	47	42	38	34	30	27	24	20	18	15	12	9	6	4
S	9	100	91	81	72	66	60	55	50	46	42	38	34	31	27	24	21	17	14	11	8	5
E	10	100	92	84	78	72	67	62	58	54	49	45	41	37	34	29	25	22	18	14	10	6
R	11	100	96	90	84	80	75	70	66	61	56	52	47	42	38	33	29	24	20	15	12	7
V	12	100	95	90	86	82	77	73	69	65	61	56	51	47	42	37	32	28	23	19	14	9
E	13	100	96	92	88	84	80	77	73	69	65	61	56	51	47	42	37	31	26	21	16	9
D	14	100	97	93	89	86	83	80	76	71	67	64	59	55	50	44	40	34	29	23	17	10
	15	100	97	93	89	86	83	80	77	74	70	66	62	57	53	47	42	36	31	24	19	11
	16	100	97	94	91	87	84	81	79	76	73	70	66	62	57	53	48	43	37	30	22	14
	17	100	98	95	92	89	87	84	82	79	77	73	70	66	62	57	52	45	39	31	24	15
	18	100	98	95	93	92	90	87	85	83	80	77	74	71	67	63	58	51	44	36	27	18
	19	100	98	96	95	93	91	89	87	85	83	80	78	75	72	68	64	58	52	44	35	24
	20	100	97	95	93	90	89	87	85	83	81	79	76	73	70	67	62	57	51	43	34	21
	21	100	99	97	96	95	94	93	92	91	90	89	87	86	84	81	78	75	69	63	55	42

Total Observations: 2405600
Total Forecasts: 2405600
Tetrachoric Slice Pts (I,J): 11 10
Tetrachoric Rgns A, B: 943600 253410
C, D: 242729 965861
Tetrachoric Fcst-Obs Corr: 0.79736

Mean Obs E.N.D.: 0.00297
Mean Fcst E.N.D.: -0.01455

Std Dev Obs E.N.D.: 0.76312
Std Dev Fcst E.N.D: 0.91753

Pearson Product-Moment Corr of E.N.D.s: 0.67019

A second, more desirable solution would be to mathematically incorporate each grid point's cloud cover climatology into the selection of its synthetic forecast cloud cover values. This should be done so that the aggregate of forecasts for all grid points would reproduce the cumulative, row-normalized SAVDOX skill matrices after a long run of the model. Since the shapes of the conditional forecast distributions depend so strongly on the distribution of observations, perhaps a technique could be devised that--for each grid point separately--biases the forecast equivalent normal deviate field by an amount proportional to the difference between the cloud cover climatology at that grid point and the hemispheric cloud cover climatology. Another way of approaching the problem might involve performing the transnormalization of input observations on an individual grid point basis, based on the climatology of that grid point or the departure of the grid point's climatology from a hemisphere or global norm. Methods of this sort would probably produce the most realistic simulation, could probably work with the present hemispheric skill matrices, and should not produce artificial discontinuities in the forecast cloud cover fields. The development of such a method, however, is beyond the scope of the present cloud forecast simulation project.

Chapter 5

SPATIAL CORRELATION IN THE CLOUD FORECAST SIMULATION MODEL

5.1 Introduction.

A design goal of the Cloud Forecast Simulation Model is to produce synthetic total cloud cover forecast fields whose spatial correlations are similar to those of the operational 3-hour cloud prognoses produced by AFGWC from the SAVDOX data base. Native SAVDOX fields exist at AFGWC 1/8-mesh resolution, but current practice is to smooth products generated from SAVDOX to 1/4-mesh resolution by application of a 9-point, 4-2-1 smoother. The present goal, then, is to simulate the spatial correlation of smoothed SAVDOX cloud prognoses.

Since AFGWC does not save the output from SAVDOX, direct studies of the spatial correlation of smoothed SAVDOX are impossible. A method of estimating or approximating the spatial correlation of smoothed SAVDOX is needed.

If it is disregarded that SAVDOX is a composite of three different models operating at different resolutions, if it is ignored that SAVDOX treats the tropics and the Southern Hemisphere differently from the Northern Hemisphere, and if it is kept in mind that the model is simulating forecasts that are only 3 hours removed in time from the 3DNEPH fields used to initialize these forecasts, then a case can be made for modeling the spatial correlation of smoothed SAVDOX forecasts on the basis of the spatial correlation of smoothed 3DNEPH fields, of which USAFETAC has an ample supply. Smoothed 3DNEPH will not have

exactly the same resolution as smoothed SAVDOX, but the spatial correlation functions should be similar.

5.2 Study of Spatial Correlation of Smoothed and Unsmoothed 3DNEPH Data.

Estimating the spatial correlation of smoothed SAVDOX fields, required an extensive set of correlation studies using 3DNEPH data for selected 3DNEPH boxes and months. Thirteen 3DNEPH grid boxes (numbers 12, 13, 14, 20, 21, 22, 23, 26, 28, 29, 30, 44, and 53) and four midseason months (January, April, July, and October) were studied, for the period 1971 through 1978. For a given box and month, the procedure was to read a field of 3DNEPH data, subject the field to the same 9-point, 4-2-1 smoother used to generate SAVDOX forecasts, and then accumulate statistics. Spatial correlations and probability distributions (see Appendix A for a description of the method used to calculate the spatial correlation of 3DNEPH data) were derived from these statistics. The 4-2-1 smoother could be turned on or off in these studies. The form of that smoother is shown in Figure 21. The spatial correlation of selected 1/4-mesh grid points with the center point of the box being studied was then calculated by the method of tetrachoric correlation. Arranging the results by distance showed how spatial correlation decays with distance. That empirical decay curve was then fitted to the Gringorten Model-B correlation function by determining the Gringorten scale distance which minimizes the root-mean-square (RMS) difference between the empirical correlation figures and the Gringorten Model-B correlations. The correlation and probability distribution calculations were made for both smoothed and unsmoothed 3DNEPH data so that the effect of the 4-2-1 smoother could be determined.

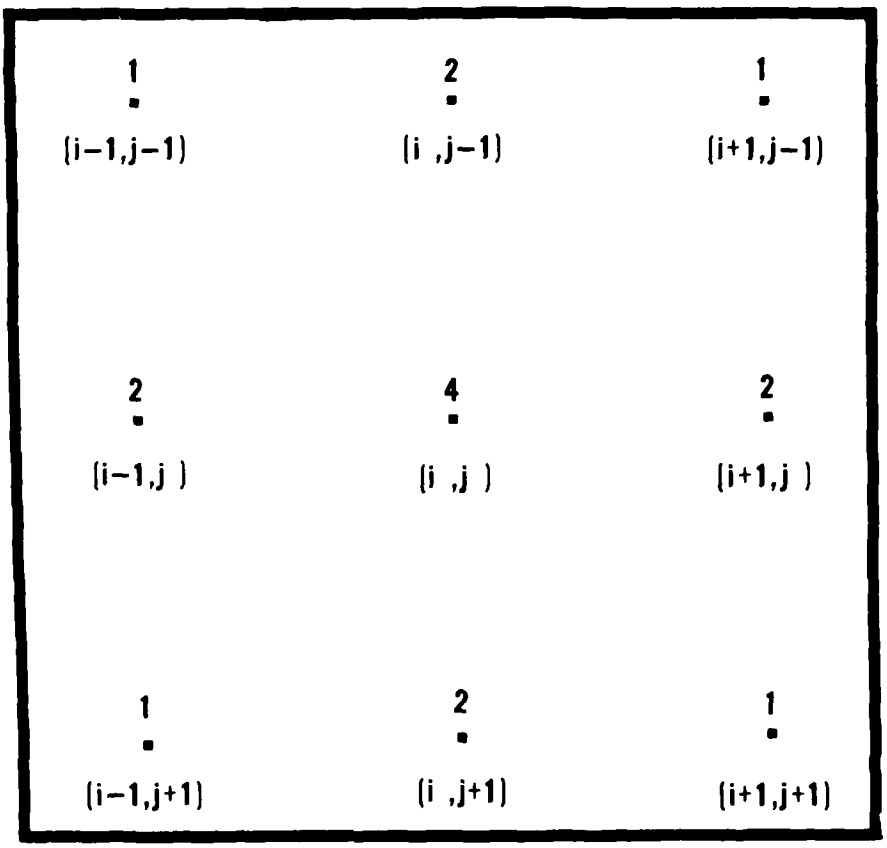


Figure 21. Weighting Factors for the 9-point 4-2-1 Smoothing Function Applied to Reduce 1/8-Mesh 3DNEPH Data to 1/4-Mesh Resolution. The value at point (i,j) is replaced by the weighted average of all 9 points, using the weighting factors shown. The smoother can still be applied if raw data for any of the 9 points is missing, by dropping those terms from the computation.

An example of how smoothing affects probability distributions is shown in Figure 22 for 3DNEPH grid box 22, which includes much of Iran and the Persian Gulf. The figure shows relative frequency distributions for smoothed and unsmoothed data. Characteristically, application of the smoother alters the relative frequency distributions by decreasing the number of clear and cloudy

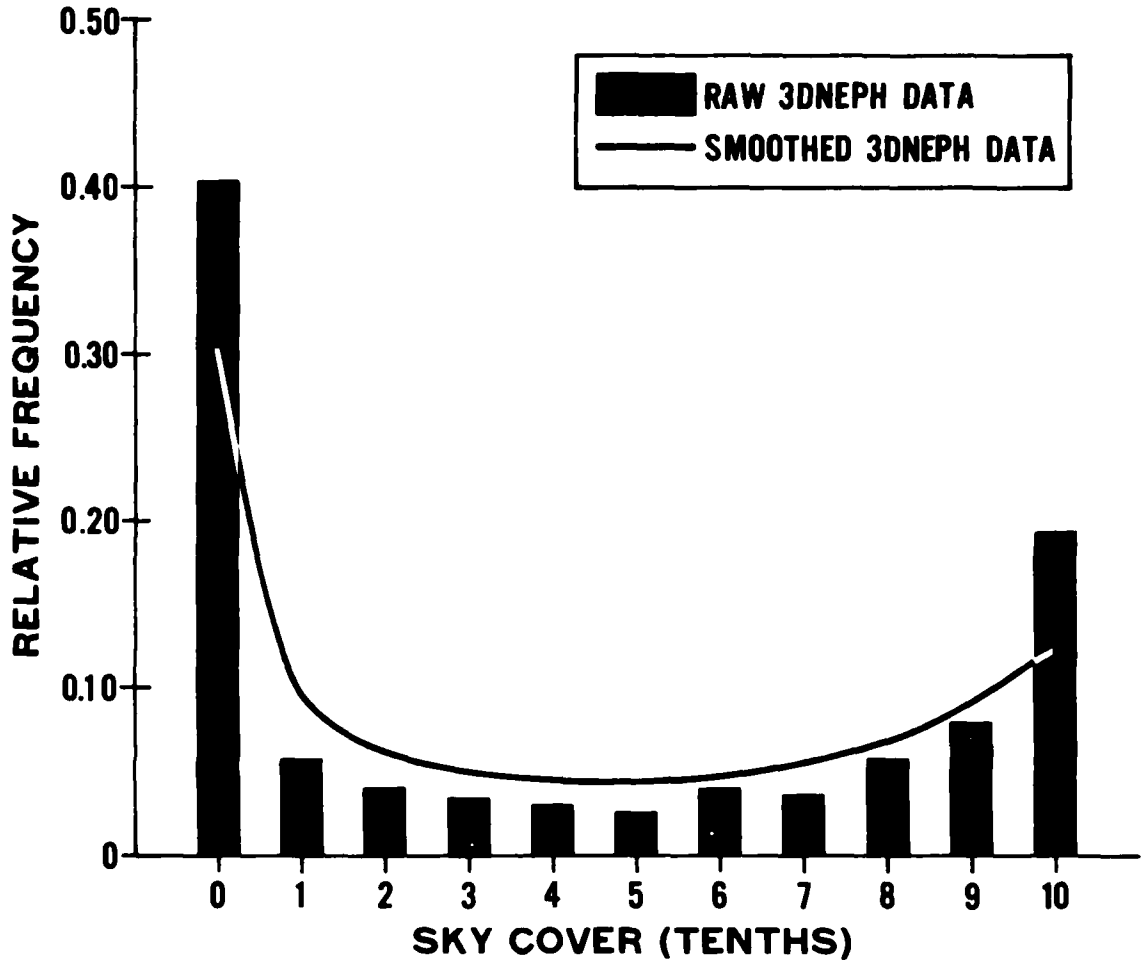


Figure 22. Relative Frequency Distributions of Cloud Cover for Raw and Smoothed 3DNEPH Data for Box 22, January, 12 LST.

cases and filling in the middle of the distribution. In this case, the relative frequency of clear is decreased 10 percentage points, from 0.41 to 0.31, while the relative frequency of overcast is decreased from 0.19 to 0.12, or 7 percentage points. The mean cloud cover for unsmoothed data is 0.37, as is that for smoothed data. The median cloud cover for the unsmoothed data is 0.30 and is 0.36 for the smoothed data. A corresponding picture for the month of July is shown in Figure 23. Here smoothing reduces the relative frequency of

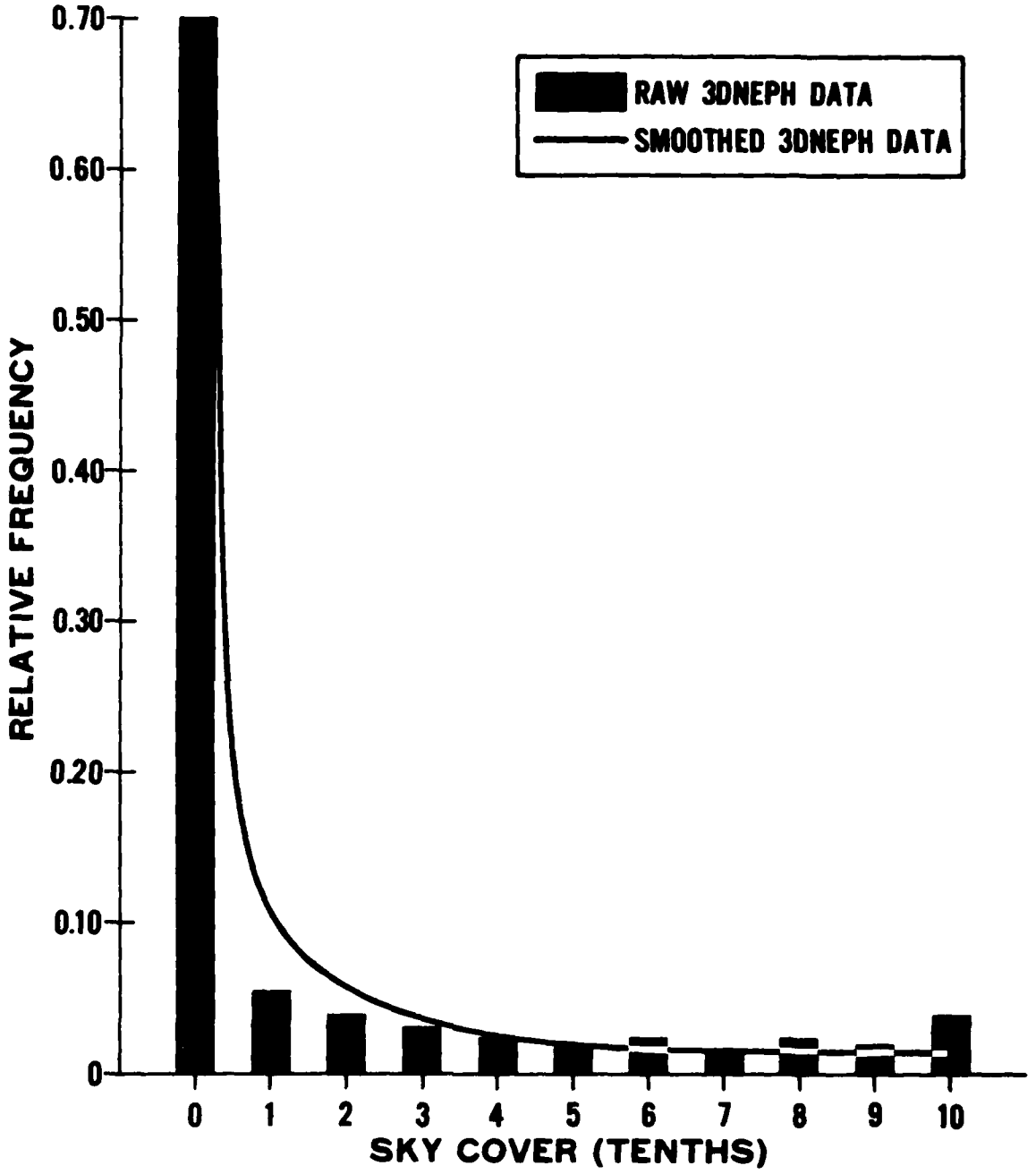


Figure 23. Relative Frequency Distributions of Cloud Cover for Raw and Smoothed 3DNEPH Data for Box 22, July, 12 LST.

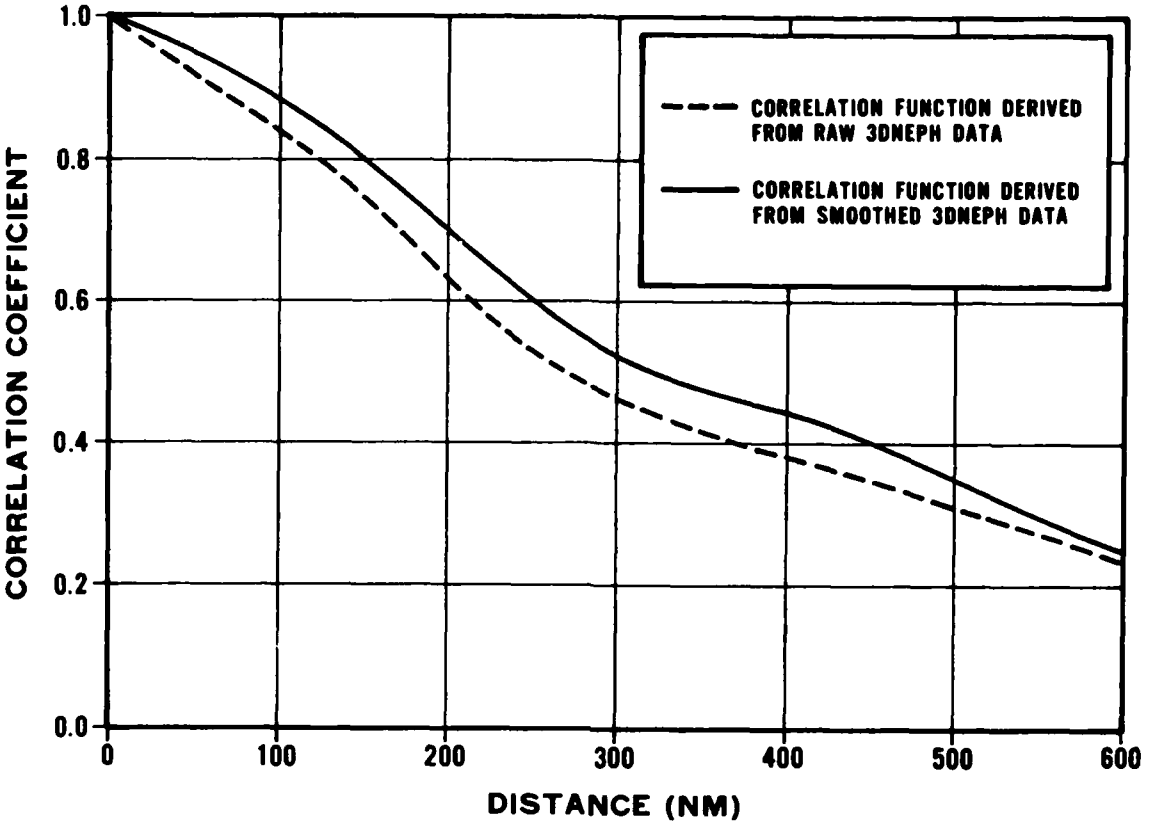


Figure 24. Comparison of Spatial Correlation Functions Derived from Raw and Smoothed 3DNEPH Data for Box 22, January. The correlation functions illustrate the typical effect of smoothing, that is, smoothing slightly increases the spatial correlation coefficients.

clear from 0.70 to 0.65, 5 percentage points, and of overcast from 0.04 to 0.02, 2 points. The mean cloud cover in the unsmoothed case is 0.13, while in the smoothed case it is 0.12. The lesson to be learned from these figures is that the probability distributions of smoothed fields (whether observations or forecasts) differ from the probability distributions of unsmoothed fields. These differences must be taken into account in designing a simulation model.

Examples of the spatial correlations calculated from 3DNEPH are shown in Figures 24 (box 22, January) and 25 (box 29, January). These curves are plots

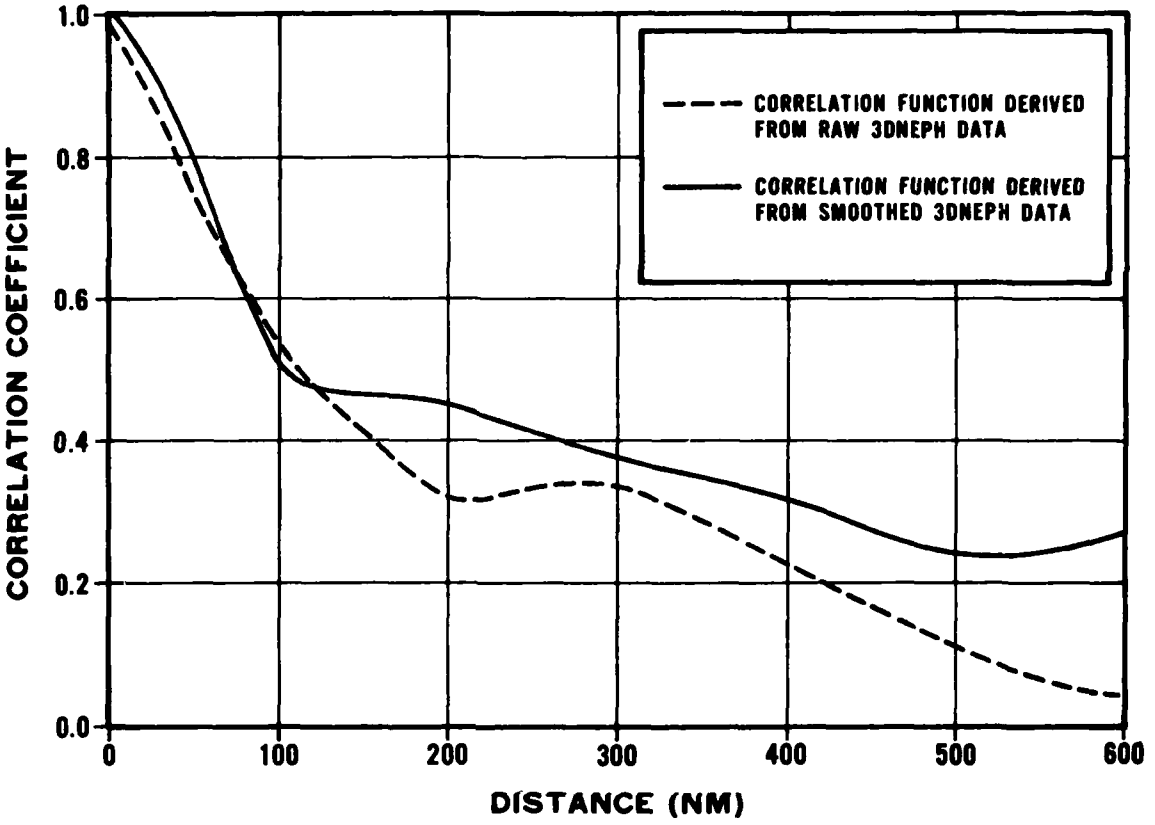


Figure 25. Comparison of Spatial Correlation Functions Derived from Raw and Smoothed 3DNEPH Data for Box 29, January. The correlation functions illustrate the typical effect of smoothing, that is, smoothing slightly increases the spatial correlation coefficients.

of spatial correlation as a function of distance. In each, the smoothed data are more highly correlated in space than the unsmoothed data at almost all distances. This effect was seen in 34 out of the 52 cases studied (65 percent) and is what intuitively might have been expected. In three cases (6 percent) the smoother had little or no effect on the spatial correlation functions, and in the remaining cases (29 percent) smoothing actually decreased the spatial correlation of the data.

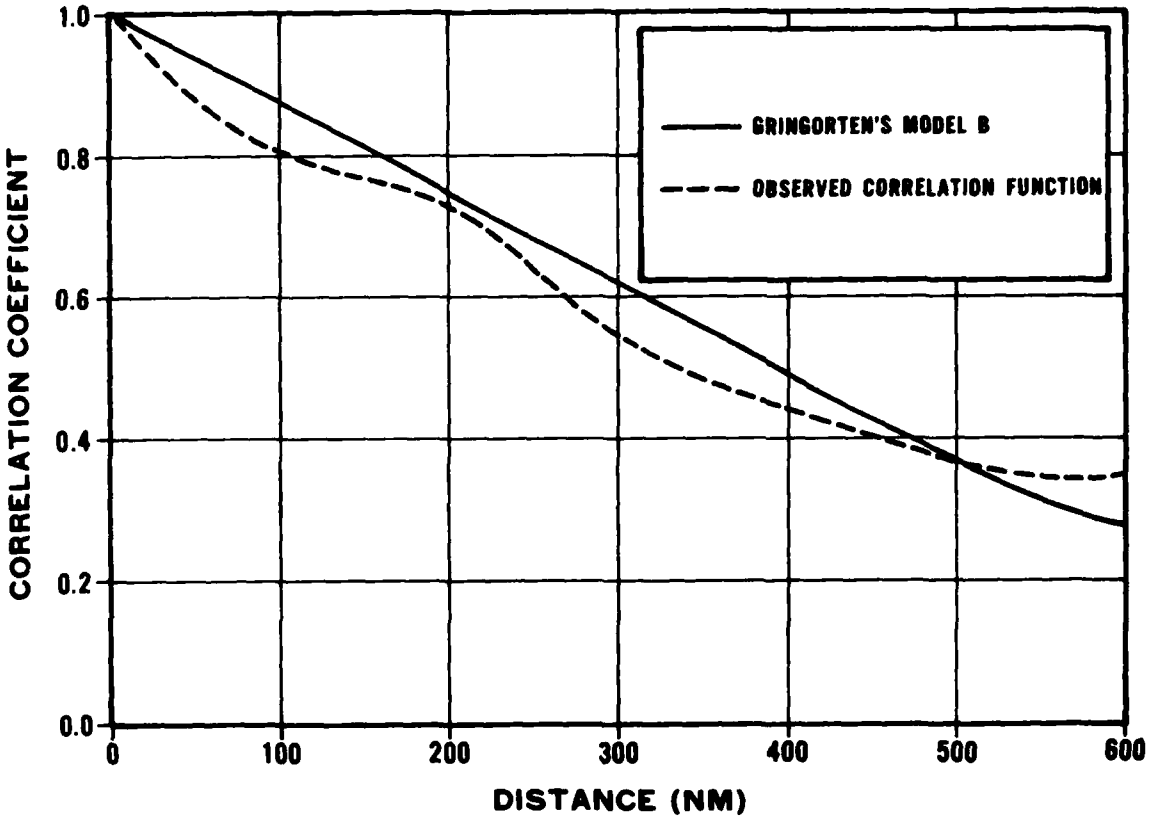


Figure 26. Comparison of the Gringorten Model-B Spatial Correlation Function for a Scale Distance of 17.0 km and a Curve Derived from Smoothed 3DNEPH Data for Box 14, January. The Gringorten curve has been fitted to the observed data by the least squares method with a RMSE of 0.05.

After spatial correlation functions were determined from the raw and smoothed 3DNEPH data for each box studied, these observed functions were used to determine characteristic Gringorten Model-B scale distances. The observed data were fit to Gringorten's Model-B so as to minimize the root-mean-square error (RMSE) over the entire range analyzed in this study (600 NM or 1100 km). Figures 26 through 28 show some typical examples of good fits. Figure 26 compares the smoothed function from box 14 (SW India and the Western Indian

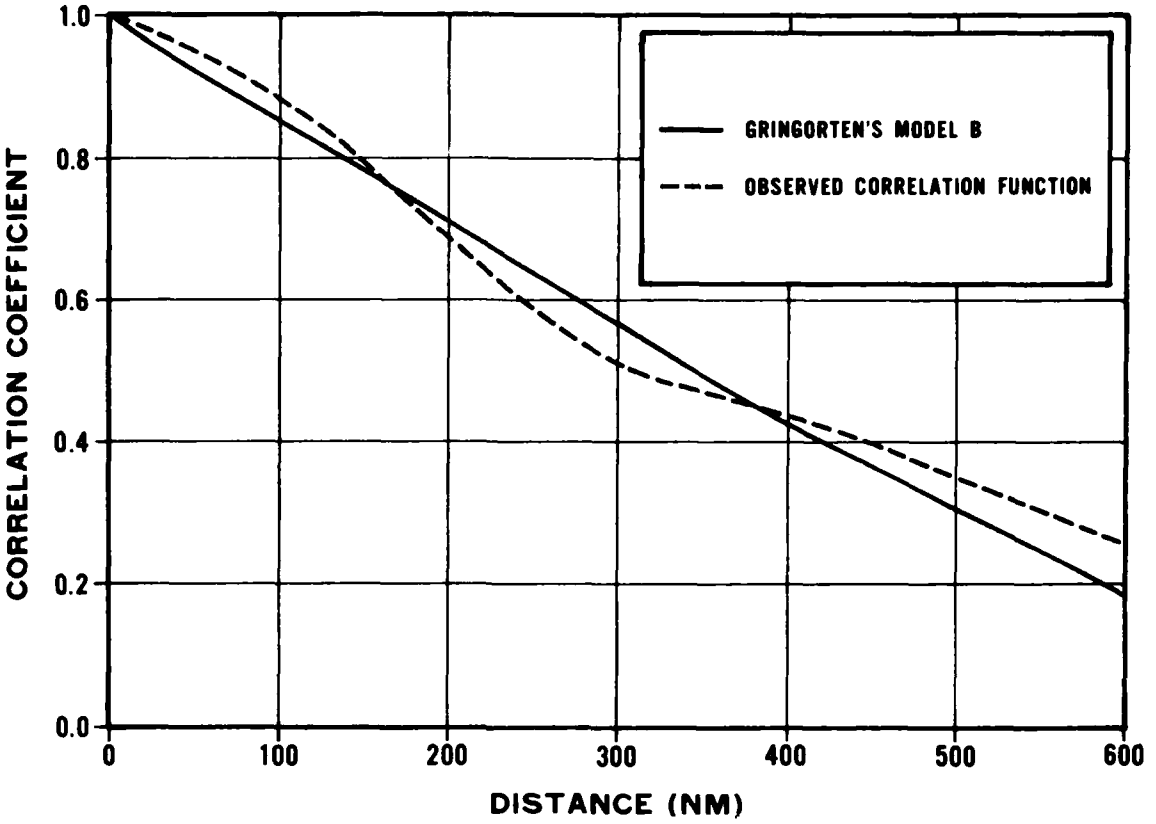


Figure 27. Comparison of the Gringorten Model-B Spatial Correlation Function for a Scale Distance of 12.8 km and a Curve Derived from Smoothed 3DNEPH Data for Box 22, January. The Gringorten curve has been fitted to the observed data by the least squares method with a RMSE of 0.03.

Ocean), January, with a Gringorten curve for a scale distance of 17.0 km. The RMSE was 0.05. Figure 27 compares the smoothed function from box 22, January, to a Gringorten curve with a scale distance of 12.8 km. The RMSE was 0.04 and was one of the best overall fits. Finally, Figure 28 compares the raw correlation function from box 12 (Southeast Asia), January, to a Gringorten curve with a scale distance of 14.4 km. The RMSE was 0.08. The results of the curve fitting procedure are summarized in Tables 11 and 12. Note that although variations of scale distance within individual boxes from raw to smoothed and from

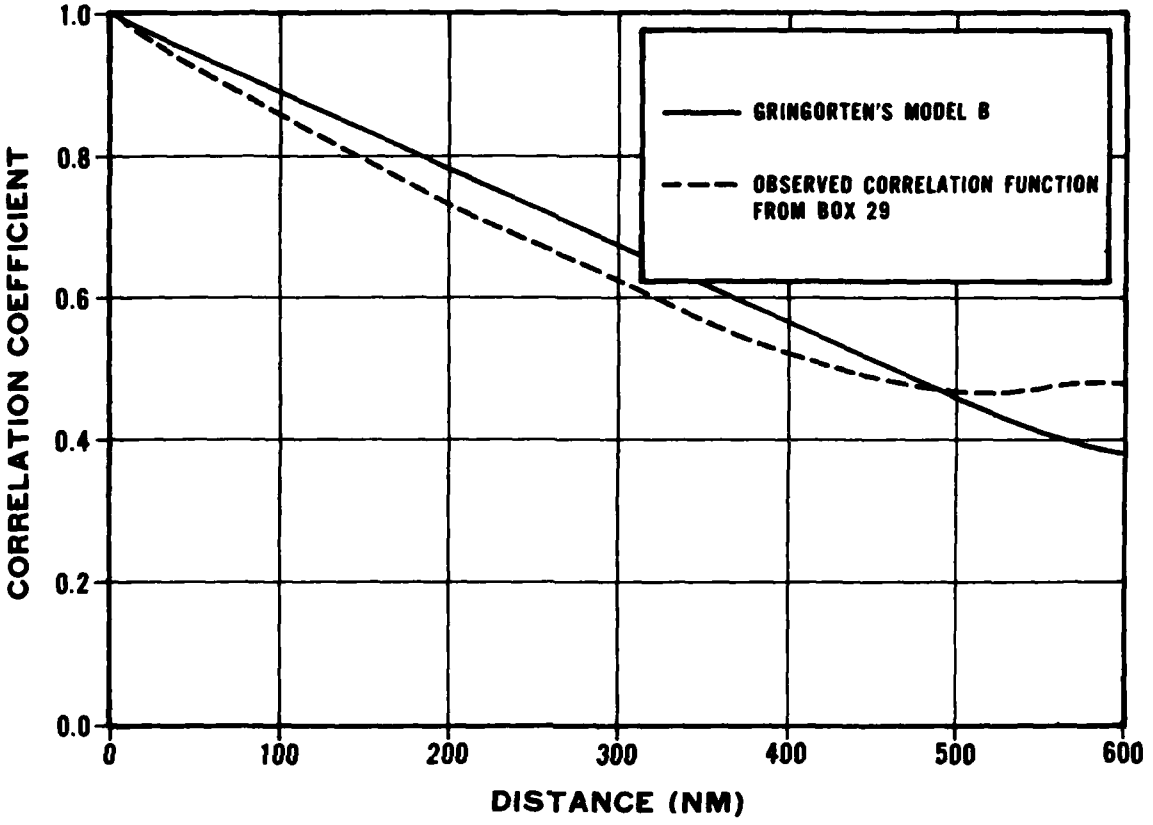


Figure 28. Comparison of the Gringorten Model-B Spatial Correlation Function for a Scale Distance of 14.4 km and a Curve Derived from Raw 3DNEPH Data for Box 12, January. The Gringorten curve has been fitted to the observed data by the least squares method with a RMSE of 0.08.

month to month are at times very large, the overall values (averages for all boxes studied) change very little from month to month and show that smoothing produces a correlation function in which correlation decreases with distance at a slower rate than is the case for unsmoothed data.

Not all box/months were as easy to fit as the previous examples. Figures 29 and 30 show some of the extreme cases. Figure 29 compares the raw and

Table 11. Gringorten's Model-B Scale Distances Derived from Raw 3DNEPH Data for Various Boxes and Months. The scale distances were obtained by fitting the Gringorten curve to the observed data by the least squares method. The numbers in parentheses represent the RMSE of the individual fits.

BOX #	JAN	APR	JUL	OCT
12	14.4 (0.08)	14.1 (0.10)	13.7 (0.09)	13.3 (0.05)
13	14.8 (0.10)	10.7 (0.17)	5.2 (0.11)	10.7 (0.10)
14	12.2 (0.09)	11.1 (0.10)	10.0 (0.13)	11.1 (0.10)
20	9.6 (0.11)	6.3 (0.07)	5.6 (0.11)	9.6 (0.05)
21	7.4 (0.09)	9.6 (0.04)	8.1 (0.11)	13.0 (0.08)
22	11.5 (0.05)	10.4 (0.06)	10.7 (0.09)	11.5 (0.10)
23	10.0 (0.16)	7.4 (0.16)	3.7 (0.09)	9.3 (0.15)
26	9.3 (0.15)	9.6 (0.14)	10.4 (0.17)	7.4 (0.17)
28	7.8 (0.11)	9.3 (0.11)	6.3 (0.07)	6.7 (0.09)
29	7.4 (0.14)	4.1 (0.06)	10.0 (0.08)	11.5 (0.19)
30	6.7 (0.17)	9.6 (0.05)	8.9 (0.14)	6.7 (0.15)
44	7.4 (0.07)	10.0 (0.03)	6.3 (0.10)	9.3 (0.07)
53	7.8 (0.13)	6.7 (0.14)	4.8 (0.12)	6.3 (0.10)
Mean	9.8 (0.11)	9.1 (0.09)	8.0 (0.11)	9.8 (0.11)

All Scale Distances are in Kilometers. Divide by 1.85 to Obtain Nautical Miles.

smoothed distributions for box 26 (mid-Pacific Ocean), July, with a Gringorten curve for a scale distance of 9.6 km. The RMSE for the Gringorten vs. smoothed fit was 0.19 and was the worst fit of all cases studied. Figure 30 compares the smoothed distribution for box 28 (northeastern USSR), January, and a Gringorten curve for a scale distance of 4.8 km. The RMSE was 0.17. The Gingorten correlation function is useful in environmental simulation modeling, but it does not appear to be capable of handling certain types of spatial correlation functions. A recurring problem with Gringorten's Model-B is its tendency to decay to zero correlation with distance more rapidly than does the correlation of the 3DNEPH data.

Table 12. Gringorten's Model-B Scale Distances Derived from Smoothed 3DNEPH Data for Various Boxes and Months. The scale distances were obtained by fitting the Gringorten curve to the observed data by the least squares method. The numbers in parentheses represent the RMSE of the individual fits.

BOX #	JAN	APR	JUL	OCT
12	15.2 (0.06)	14.8 (0.08)	15.5 (0.10)	20.7 (0.08)
13	12.2 (0.12)	10.7 (0.16)	11.5 (0.18)	10.7 (0.16)
14	7.0 (0.05)	14.4 (0.07)	13.7 (0.08)	11.5 (0.11)
20	8.9 (0.09)	7.0 (0.10)	13.7 (0.16)	10.0 (0.04)
21	6.3 (0.09)	9.3 (0.06)	13.7 (0.08)	10.0 (0.09)
22	12.8 (0.04)	11.1 (0.05)	11.5 (0.06)	11.1 (0.07)
23	10.4 (0.16)	8.5 (0.13)	4.1 (0.14)	5.6 (0.14)
26	9.6 (0.12)	10.4 (0.18)	9.6 (0.19)	8.9 (0.13)
28	4.8 (0.17)	7.4 (0.17)	10.0 (0.06)	11.1 (0.07)
29	9.6 (0.16)	9.6 (0.16)	7.4 (0.09)	19.2 (0.12)
30	12.2 (0.15)	14.4 (0.05)	10.0 (0.10)	5.9 (0.16)
44	7.4 (0.10)	12.2 (0.07)	10.4 (0.09)	9.6 (0.11)
53	6.7 (0.14)	6.3 (0.13)	8.5 (0.17)	4.8 (0.14)
Mean	10.2 (0.11)	10.5 (0.11)	10.7 (0.12)	10.7 (0.11)

All Scale Distances are in Kilometers. Divide by 1.85 to Obtain Nautical Miles

Comparing Tables 11 and 12 shows that smoothing does not always increase the spatial correlation of 3DNEPH data. In box 26 during July, for example, the Gringorten scale distance for unsmoothed data was 5.6 NM (10.4 km), while that for smoothed data was 5.2 NM (9.6 km). The greater spatial correlation in this case is illustrated by Figure 29. Note also that Gringorten's Model-B did not fit this difficult box/month. RMS errors were among the highest obtained in all the box/months fitted. Another case in which the spatial correlation of unsmoothed data (8.0 NM or 14.8 km) was greater than that (6.6 NM or 12.2 km) of smoothed data is illustrated in Figure 31, showing box 13 in January. Of the 52 box/months analyzed for this study, 15 (28 percent) exhibited this sort

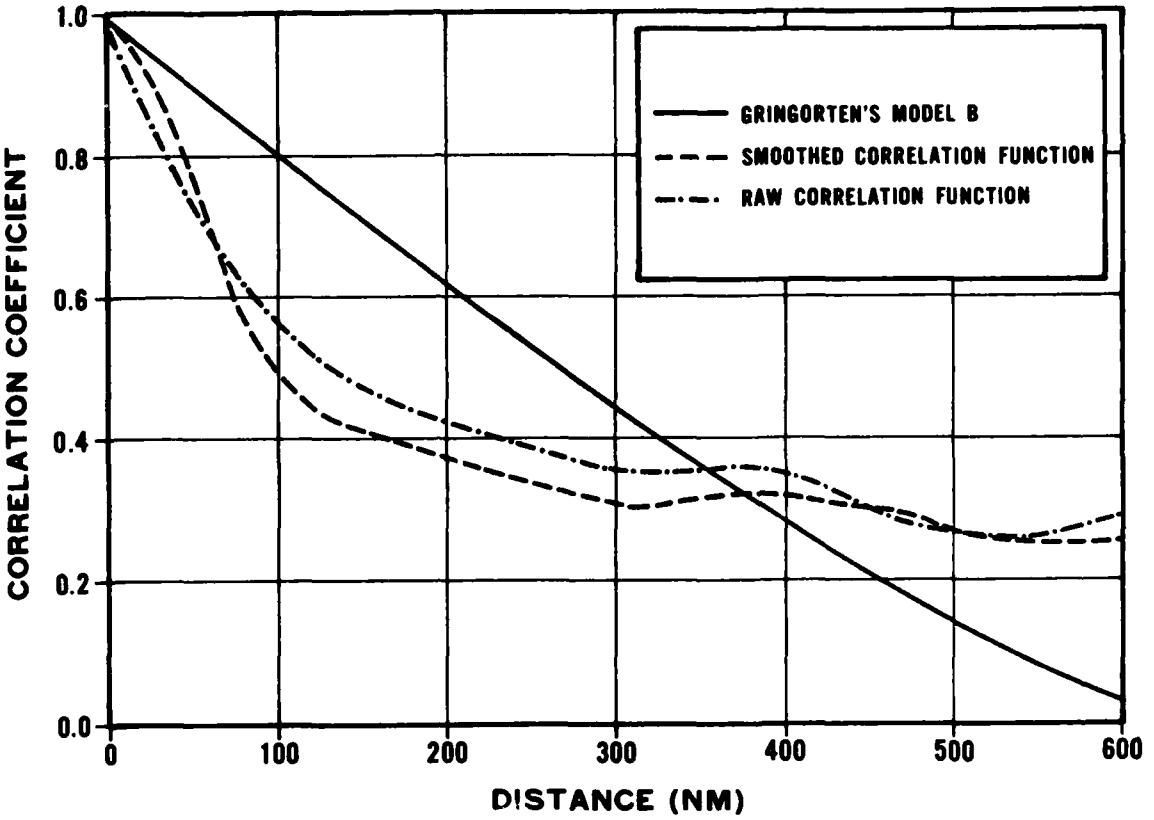


Figure 29. Comparison of the Gringorten Model-B Spatial Correlation Function for a Scale Distance of 9.6 km and Curves Derived from Raw and Smoothed 3DNEPH Data for Box 26, July. The Gringorten curve has been fitted to the smoothed data by the least squares method with a RMSE of 0.19.

of anomaly. Fully 67 percent of these abnormalities occurred in January and October (four months were studied), and 53 percent occurred in 3DNEPH grid boxes 21, 28, and 53 (13 boxes were studied). Box 28 includes the Arctic snow and ice pack, which can invalidate and distort 3DNEPH data. The difficulty with box 21, which covers the central Sino-Soviet, is hard to explain but may be caused by the snow cover and the discontinuities of cloud cover due to the rugged terrain of the Himalayas. See Table 13 for a full breakdown of these anomalies by box and month.

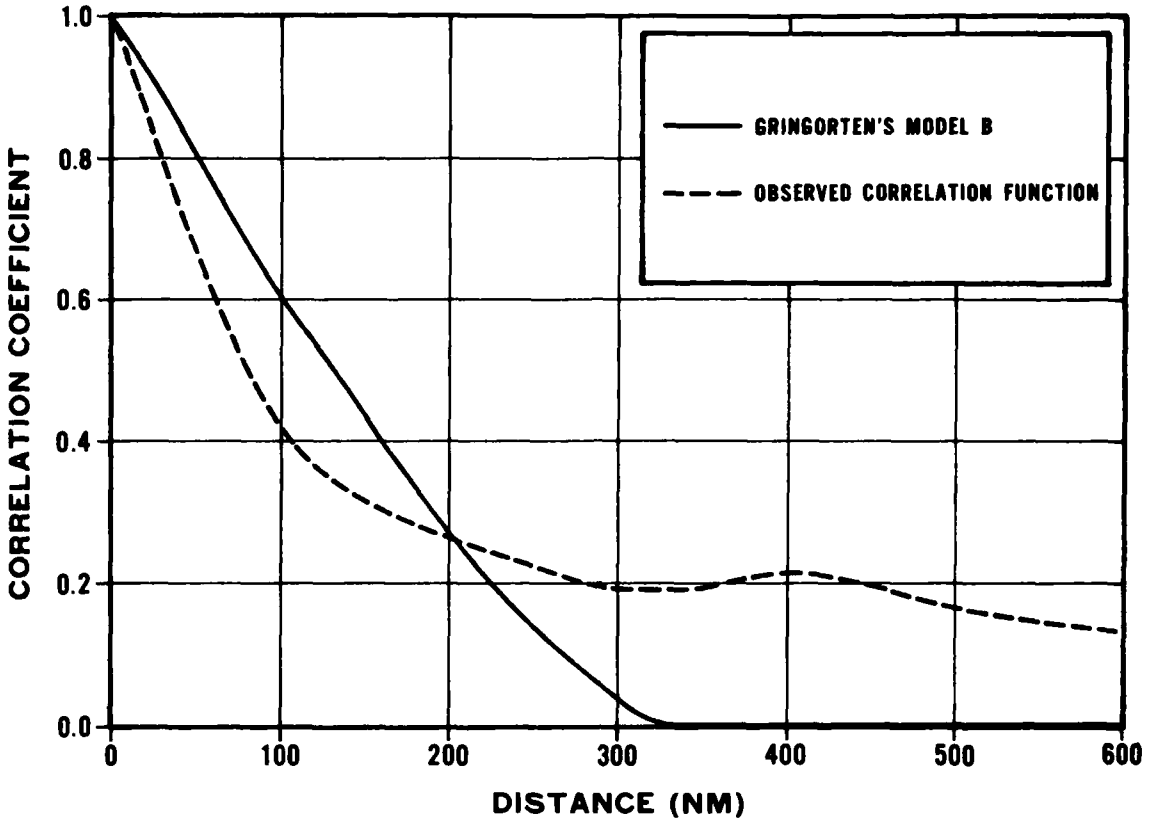


Figure 30. Comparison of the Gringorten Model-B Spatial Correlation Function for a Scale Distance of 6.8 km with a Curve Derived from Smoothed 3DNEPH Data for Box 28, January. The Gringorten curve has been fitted to the observed data by the least squares method with a RMSE of 0.17.

Gringorten's Model-B could be fitted to raw 3DNEPH spatial correlation with an RMSE less than or equal to 0.15 in 45 of the 52 cases studied (87 percent). Model-B could be fitted to smoothed data within the same RMSE in 40 of the 52 cases (77 percent). Mean RMSEs for fits to unsmoothed 3DNEPH were slightly less than the mean RMSEs for fits to smoothed 3DNEPH. Two reasons could account for this. First, smoothing usually increases spatial correlation, so the spatial correlation function of smoothed 3DNEPH data is harder to fit with the

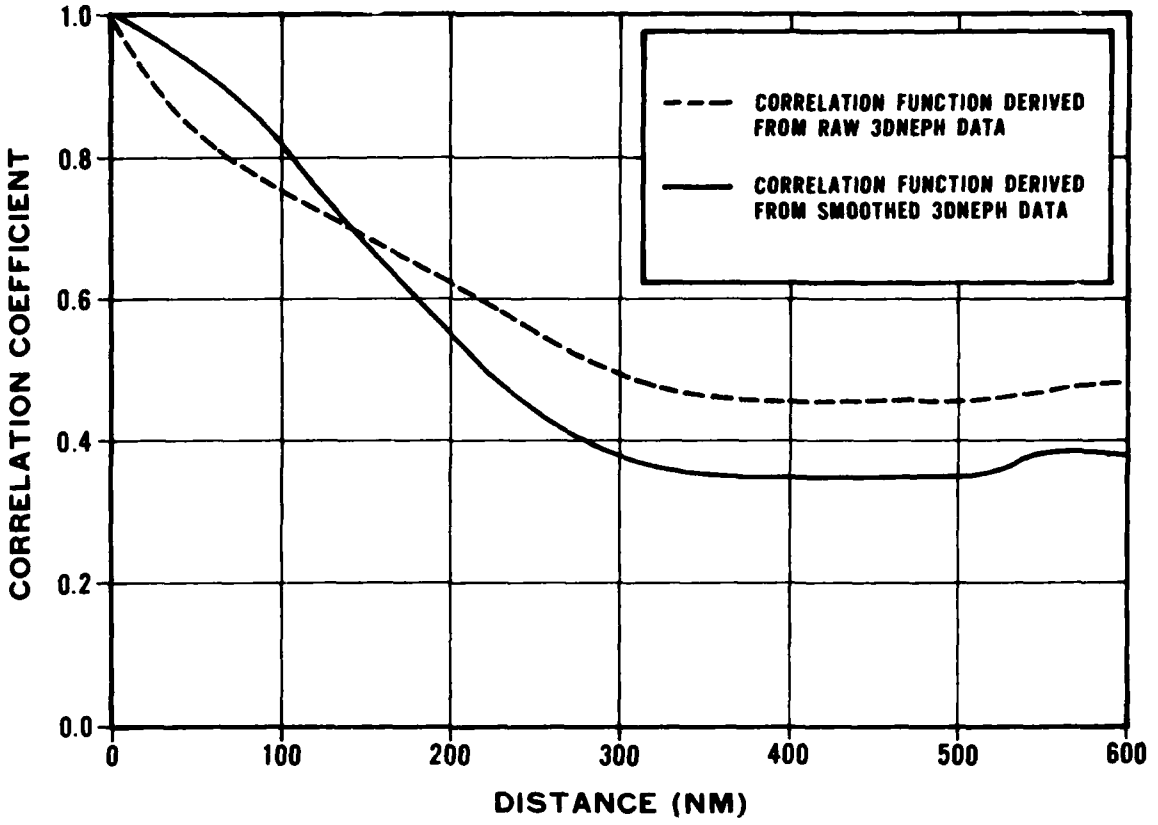
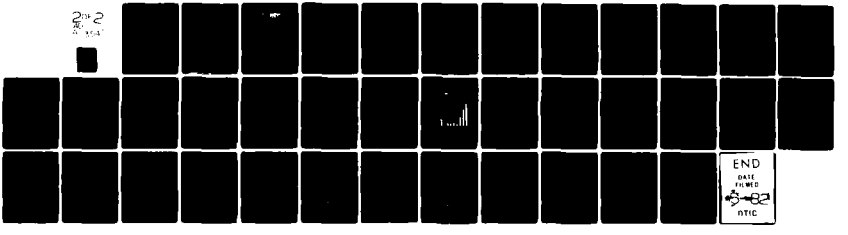


Figure 31. Comparison of the Spatial Correlation Functions Derived from Smoothed and Raw 3DNEPH for Box 13, January. The correlation function illustrates that in a small number of cases smoothing increases the spatial correlation coefficients in the first 0-100 NM but may act to reduce the coefficients beyond 100 NM.

Gringorten function, which decreases rapidly to zero. Second, Gringorten developed his function from surface data, so it is not surprising that the raw 3DNEPH data, which is heavily influenced by surface observations, can be better described by Model-B than the smoothed 3DNEPH data.

AD-A113 540 AIR FORCE ENVIRONMENTAL TECHNICAL APPLICATIONS CENTER--ETC F/G 4/2
CLOUD FORECAST SIMULATION MODEL.(U)
OCT 81 R C WHITON, E M BERECEK, J C SLADEN
UNCLASSIFIED USAFETAC/TN-81/004 SBI-AD-E850 149 NL



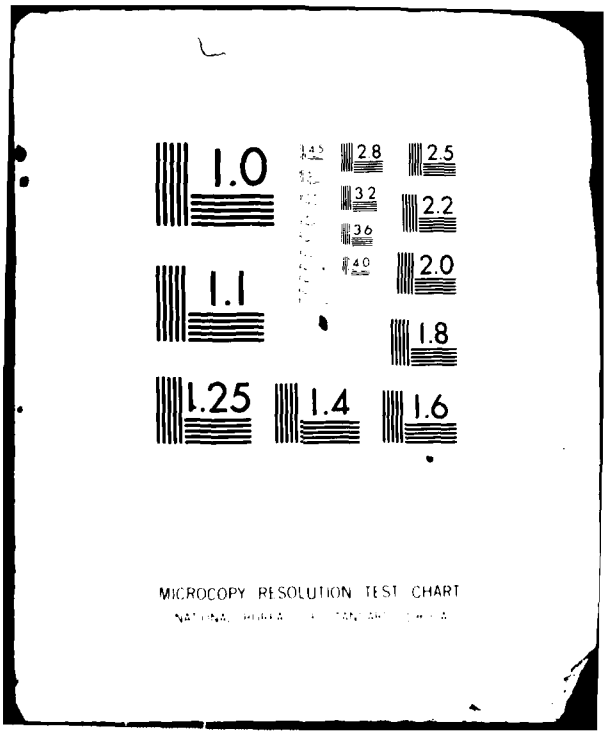

1.0
1.1
1.25
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

Table 13. Breakdown by Month and Box of the Cases in Which Smoothing of 3DNEPH Data Decreased Spatial Correlation.

Box #	Jan	Apr	Jul	Oct	Total
13	X				1
20	X				1
21	X	X		X	3
22				X	1
23				X	1
26			X		1
28	X	X			2
29			X		1
30				X	1
53	X	X		X	3
Total	5	3	2	5	15

5.3 Study of Spatial Correlation Within the Cloud Forecast Simulation Model (FCLDO).

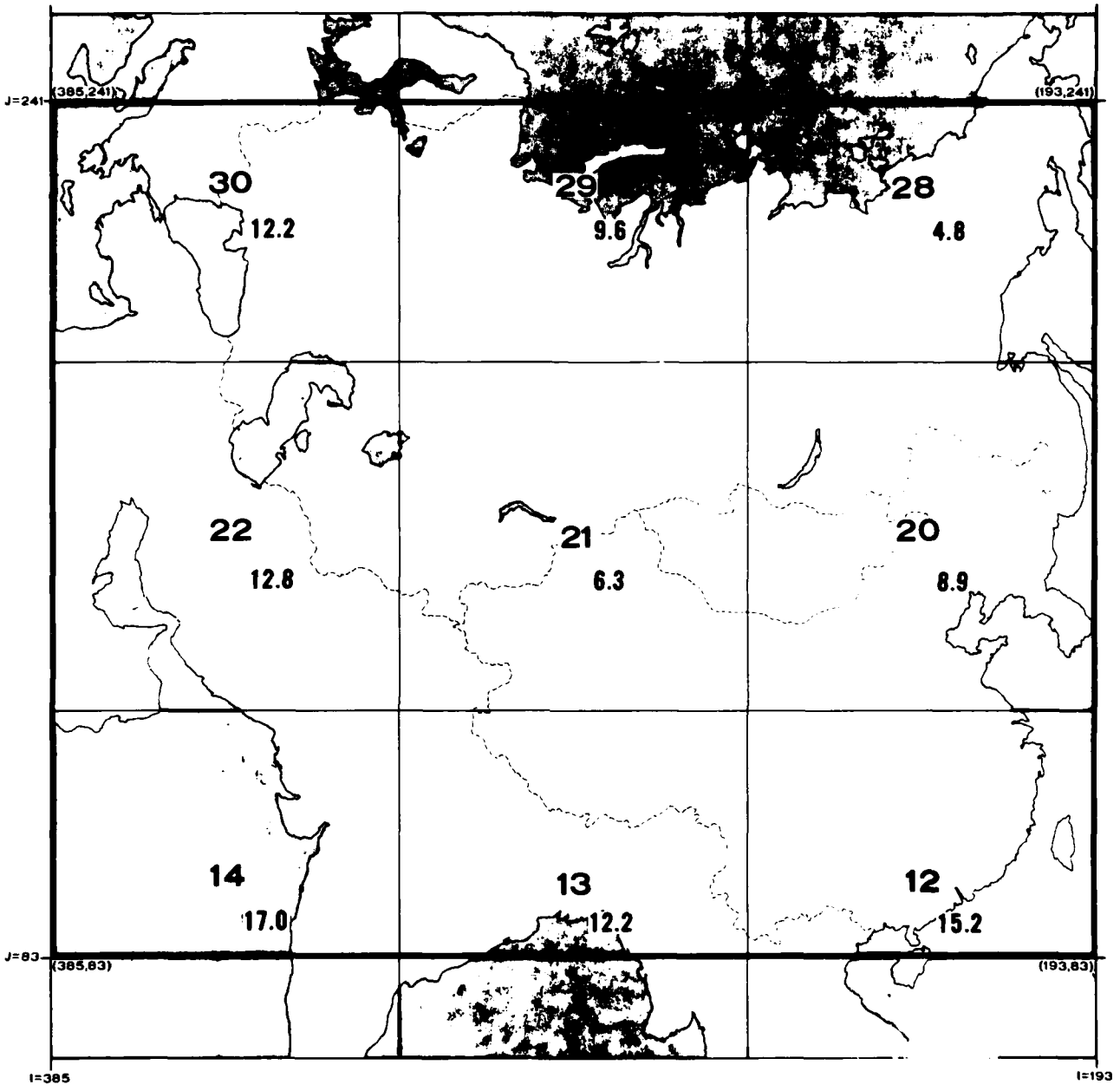
In the previous section of this report, results of the study of the spatial correlation of 3DNEPH data are reported. Now the question is, how do the spatial correlation functions of the Cloud Forecast Simulation Model's (FCLDO) synthetic forecasts match that of smoothed 3DNEPH? First, consider how the spatial correlation functions of FCLDO were derived.

The spatial correlation functions of three different fields of FCLDO were of interest: (1) the random normal number fields (η) produced by the sawtooth

wave submodel, (2) the fields of ENDs of the synthetic forecasts ($\underline{\underline{e}}_f$) produced by the cloud forecast simulation equation, and (3) the fields of ENDs of the input observations ($\underline{\underline{e}}_o$) used to generate the synthetic forecasts. Correlation data were accumulated for each. Although the sawtooth wave submodel of FCLDO could operate on a hemispherical or global scale, early development and testing were performed on a 97 x 80 subgrid of the AFGWC 3DNEPH supergrid system (Figure 32 outlines this developmental subgrid). The correlation calculations proceeded in three steps:

(1) Pre-simulation preparations: Tetrachoric tables as described in Appendix A, paragraph d, were set up for each 50-NM increment, up to 600 NM, for the three fields being studied. Eight reference points were chosen at random, from different portions of the subgrid. The only limitation on the selection of these points was that they must have been at least 13 grid points (600 NM) in from the edge of the subgrid. This limitation prevented the grid pair selection pattern from going outside the subgrid.

(2) As the simulation proceeded: On each forecast replication of FCLDO, the individual fields' tetrachoric table for each distance was incremented by using the data at each reference point and the data at selected grid points as data pairs. Figure A-1 of Appendix A illustrates the selection pattern from each reference point. The two range marks represent the 100 and 200 NM distance circles superimposed on a 1/4-mesh grid. For distances less than or equal to 100 NM, the two-by-two tables were incremented with all data pairs in the same relative position as the C-A grid pairs in Figure A-1. For distances 150 to 600 NM, additional points were added, and the tables were incremented with all data pairs that are in the same relative position as the C-B



STRAIGHT AVERAGE OVER ENTIRE SUBGRID: 10.9
WEIGHTED AVERAGE: 11.5

Figure 32. Gringorten Model-B Scale Distances for Individual Boxes in the 97 x 80 Subgrid. The scale distances were derived from smoothed 3DNEPH total cloud cover, January. All scale distances are in kilometers (divide by 1.85 to obtain NM).

grid pairs in this figure. In this manner, 32 correlation pairs (8 reference points x 4 selected points) were input into the tables for 50 and 100 NM, and 96 correlation pairs (8 reference points x 12 selected points) were input into the tables for 150 to 600 NM, for each replication (i.e., each synthetic forecast field, $\underline{\eta}$ field, or observation field).

(3) Simulation termination and wrapup: At the conclusion of the simulation, tetrachoric correlation coefficients were calculated from each two-by-two table using the false position algorithm described in Appendix A. The characteristic spatial correlation functions of the three fields of interest were derived from these coefficients, and the results are described in the following paragraphs.

The study of spatial correlation within FCLDO progressed in several distinct phases:

(1) During earlier work with the sawtooth wave generator, Major Al Boehm, USAFETAC/DNP, developed empirical equations that converted a desired Gringorten Model-B scale distance (SD) into maximum and minimum allowable wavelengths [W(upper) and W(lower)] for the wave generator. These equations are

$$W(\text{upper}) = C(u) * SD \tag{63}$$

$$W(\text{lower}) = C(1) * SD \tag{64}$$

The dimensionless constants C(u) and C(1) are selected so the model returns a spatial correlation function for the $\underline{\eta}$ field that has a shape similar to a

Gringorten curve for a particular scale distance SD. Selecting SD as input to the model presents a problem. The major drawback to the sawtooth wave model is that the chosen spatial correlation function is virtually constant over the entire $\underline{\eta}$ field. This is not the case in observed cloud cover data. Consider the 97 x 80 subgrid (Figure 32). The subgrid spans portions of nine 3DNEPH boxes, each exhibiting different cloud cover distributions and spatial correlation functions. Figure 32 shows the scale distances derived from smoothed, January 3DNEPH data for the individual boxes. The scale distances for the nine boxes range from 4.8 to 17.0 km. One spatial correlation function does not adequately describe the entire subgrid. This problem is compounded when the model is expanded to global scale. Since a single scale distance SD must be input into FCLDO to determine the input wavelengths, a single characteristic SD must be found for the model.

Several approaches can be taken to derive the characteristic scale distance. One way is to take a straight average over the nine different scale distances of the subgrid; this gives a resulting scale distance of 10.9 km. Another method is to use some type of weighting scheme, with high interest areas carrying more influence. For example in Figure 32, Europe, the Middle East, and the western USSR might be of more interest than SE Asia. The characteristic scale distance SD for the model can then be determined by averaging the individual scale distances for boxes 22, 29, and 30, giving a resulting scale distance of 11.5 km. Figure 33 illustrates how the Gringorten curve chosen by the weighting method provides a good middle-of-the-road fit between the spatial correlation functions of smoothed 3DNEPH data for boxes 22 and 29. There is, however, a tradeoff. Using this Gringorten curve results in a terrible fit over box 28, as shown in Figure 34. The desired objectives of the

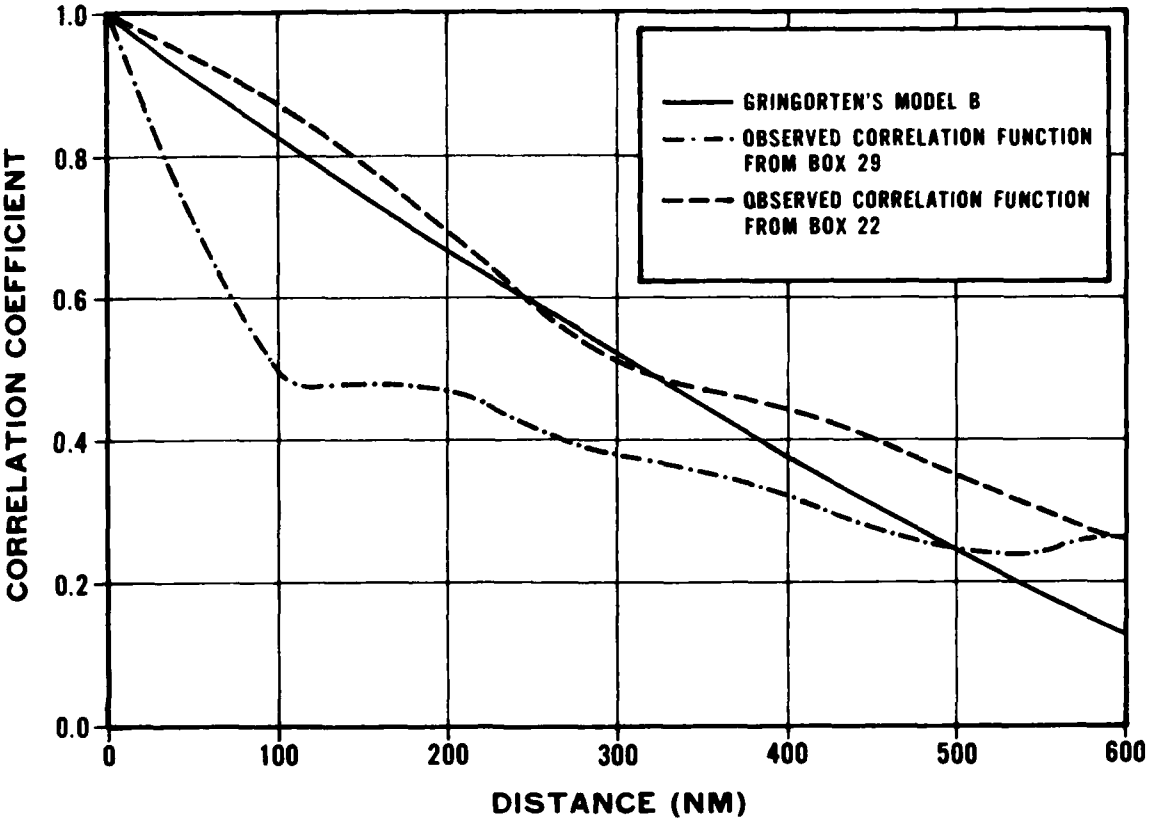


Figure 33. Comparison of the Spatial Correlation Function of Gringorten's Model-B for a Scale Distance of 11.5 km with Curves Derived from Smoothed 3DNEPH Data for Box 22 and Box 29, January. The Gringorten curve has been chosen to provide a middle-of-the-road fit between the two observed functions in an attempt to tune the cloud forecast model to certain geographical areas.

modeling effort will dictate the types of tradeoffs that can and cannot be tolerated.

Once the characteristic scale distance was found, the next step was to use it to determine the input wavelengths for the sawtooth wave submodel. Initially Boehm's recommended constants of 50 and 350 for C(1) and C(u), respectively, were used. The early tests of FCLDO showed that these constants produced a

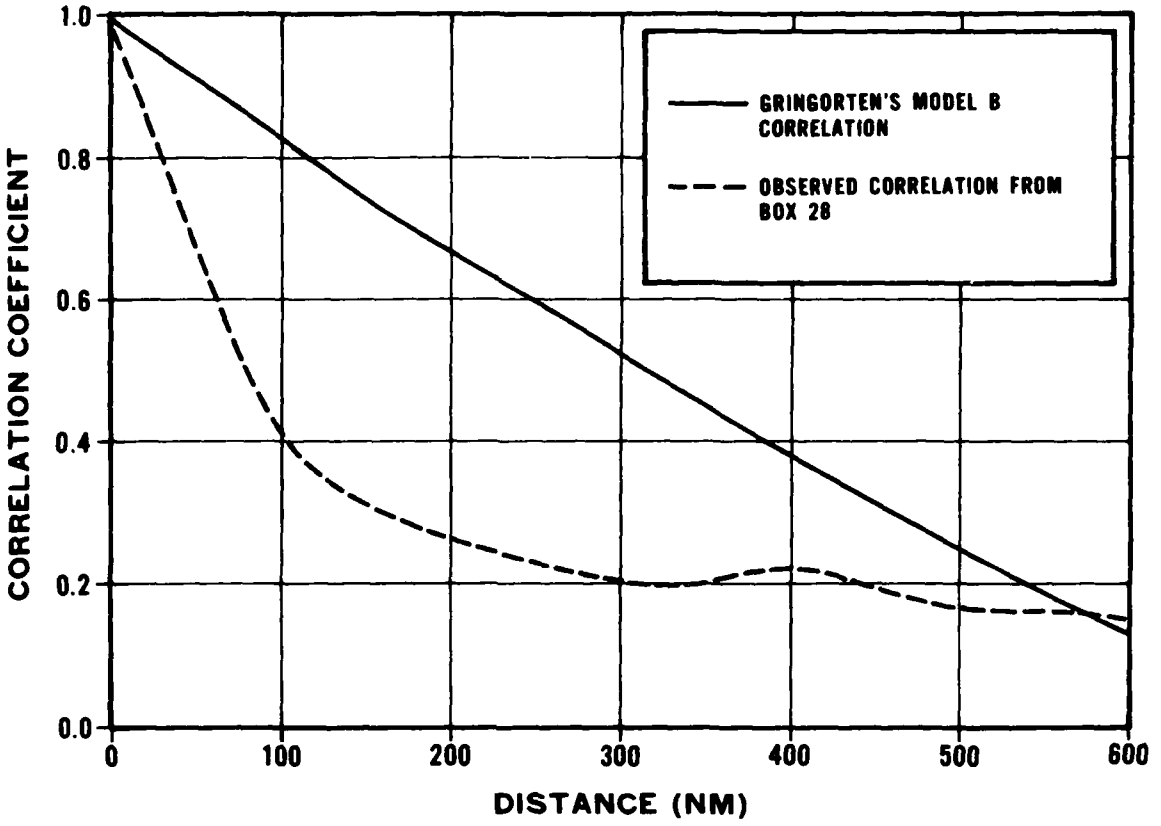


Figure 34. Comparison of the Spatial Correlation Function of Gringorten's Model-B for a Scale Distance of 11.5 km with a Curve Derived from Smoothed 3DNEPH Data for Box 28, January. Tuning the model to one area may result in bad fits in other areas.

correlation function for the $\underline{\eta}$ field that was systematically lower, or less spatially correlated, than the desired Gringorten curve. A series of tests was then made to determine new constants and remedy this situation. The resulting new values [C(l) = 175 and C(u) = 450] gave the closest match between the correlation function of the $\underline{\eta}$ field and that of the Gringorten curve.

As an example, it might be desirable for the spatial correlation function of the $\underline{\eta}$ field to resemble a Gringorten Model-B curve for a scale distance of

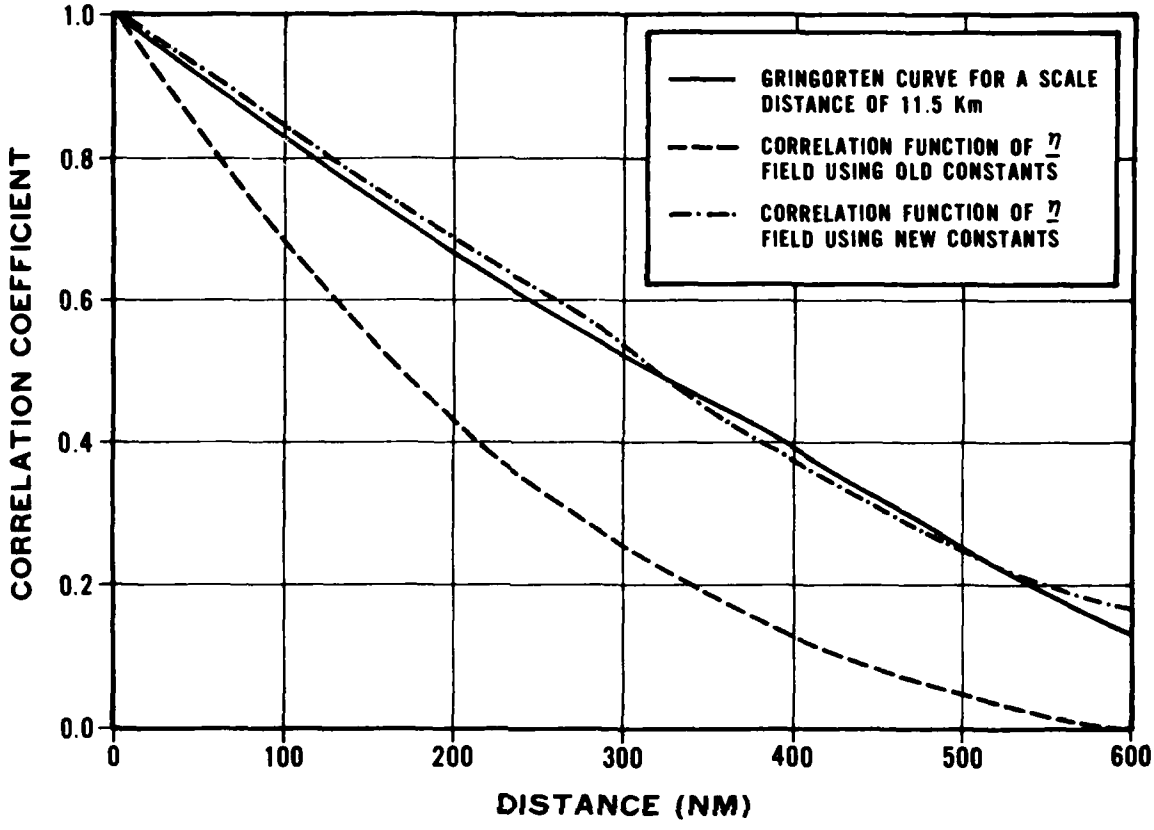


Figure 35. Comparison of the Spatial Correlation Functions Derived from FCLDO's $\underline{\eta}$ Field using the Old (C(1)=50, C(u)=350) and New Constants (C(1)=175, C(u)=450) with a Gringorten Model-B Curve for a Scale Distance of 11.5 km.

11.5 km. Substituting this scale distance into Equations (63) and (64), input wavelengths for the model are derived:

$$W(\text{upper}) = 450 * 11.5 \text{ km} = 5175.0 \text{ km}$$

$$W(\text{lower}) = 175 * 11.5 \text{ km} = 2012.5 \text{ km}$$

Figure 35 compares the spatial correlation functions of FCLDO's $\underline{\eta}$ field using the old and new constants with a Gringorten curve for a scale distance of

Table 14. Results of Sample Runs of Cloud Forecast Simulation Model Using Values of 175 and 450 for C(1) and C(u), Respectively.

SD of Desired Gringorten Curve (km)	Recommended Wavelengths W(lower) (km)	W(lower) (km)	SD Derived (by Least Squares) from Resultant $\underline{\eta}$ Correlation Function (km)	RMS of Least Squares Fit (km)
3.0	525	1350	2.96	0.007
5.0	875	2250	4.85	0.004
7.0	1225	3150	6.66	0.026
10.0	1750	4500	9.95	0.015
15.0	2625	6750	14.95	0.014
20.0	3500	9000	19.23	0.013

11.5 km. Table 14 shows the results of six different tests of the model using the new constants. In each case the $\underline{\eta}$ field's correlation function came very close to the Gringorten curve the model was attempting to duplicate. The verification of the closeness of fit between the two functions is expressed as an average absolute difference or root-mean-square (RMS) difference. Note that these RMS differences between the two functions are very small.

(2) After it was determined that the sawtooth wave submodel was satisfactorily producing a random normal number field with the desired spatial correlation, the next step was to study the spatial correlation function of the synthetic forecast field. Since the model produces synthetic forecasts by two methods, a pilot study comparing the spatial correlation functions of the basic model and the hybrid sawtooth wave/skill matrix design was conducted. The study consisted of making a pair of simulation runs using the same input observations, input wavelengths, number of forecast replications, etc., the only difference being the method used to generate the synthetic forecast fields.

Table 15 summarizes the results of two of these tests. In the eight cases studied, the correlation coefficients for the two methods did not differ by more than 0.01. The spatial correlation functions produced by the two techniques are virtually identical; however, the forecast adjustment technique proved to be more desirable because it was better able to reproduce an input skill matrix (see Chapter 4 of this report). The following discussion deals with the spatial correlation functions produced by the basic model, but could be extended to the forecast adjustment technique because of the similarity in the spatial correlation functions produced by these two methods.

Table 15. Comparison of the Spatial Correlation Functions of Synthetic Forecast Fields Produced by FCLDO's Basic Model and the Hybrid Forecast Adjustment Technique. Input observation fields were January 1979, 12 LST, MPS data.

	W(lower) = 2200 km W(upper) = 5500 km		W(lower) = 575 km W(upper) = 5025 km	
Distance (NM)	Adjusted	Basic	Adjusted	Basic
100	0.822	0.825	0.745	0.749
200	0.663	0.670	0.556	0.549
300	0.541	0.542	0.415	0.406
400	0.387	0.393	0.274	0.266
500	0.278	0.281	0.197	0.189

Consider the basic model equation, Equation (4), that gives a value for an END of a forecast at each grid point. The synthetic forecast END ($\overset{''}{c}_f$) is the sum of two parts: (a) a deterministic term derived from the END of the input observation ($\overset{''}{c}_o$) and (b) a stochastic term, embodying the imperfection of weather forecasts. The stochastic portion of this equation, in effect, decorrelates the synthetic forecast from the verifying observation. The amount of decorrelation is governed by the forecast-observation correlation (ρ_{fo}). The

spatial correlation of part (a) is governed by the unique correlation function of the input observation fields (MPS data). The spatial correlation of part (b) is governed by the correlation function of the $\underline{\eta}$ field. In theory then, the spatial correlation function of the synthetic forecasts will be a mix of that of the input observation field and that of the $\underline{\eta}$ field. For short term forecasts, where forecast-observation correlation is high, the input observations will weigh more heavily than the random normal $\underline{\eta}$ field in determining the spatial correlation of output synthetic forecasts. For long term forecasts, where forecast-observation correlation is low, the correlation function of the $\underline{\eta}$ field will dominate. When input wavelengths for the sawtooth wave generator are intentionally selected so that the spatial correlation function of the $\underline{\eta}$ field resembles that of input MPS data, the result should be that part (b) of the equation decorrelates the synthetic forecasts from the input observations but should not act to decorrelate the forecasts in space. This can be seen by considering an example.

Figure 36 shows results from a sample run of FCLDO using January 1979, 12 LST, MPS data as input observations. Ten forecasts were made from each of 31 input observation fields. The input wavelengths for the sawtooth wave generator were 2012.5 km for W(lower) and 5175.0 km for W(upper). This correlation diagram shows that the spatial correlation function of the $\underline{\eta}$ fields, the synthetic forecast fields, and the input observation fields are all very close. There appears to be only a slight decorrelation in space of the synthetic forecasts when compared to the input observations.

(3) The final phase in the testing of FCLDO dealt with fine-tuning the model. For example, if a historical record of SAVDOX forecasts becomes avail-

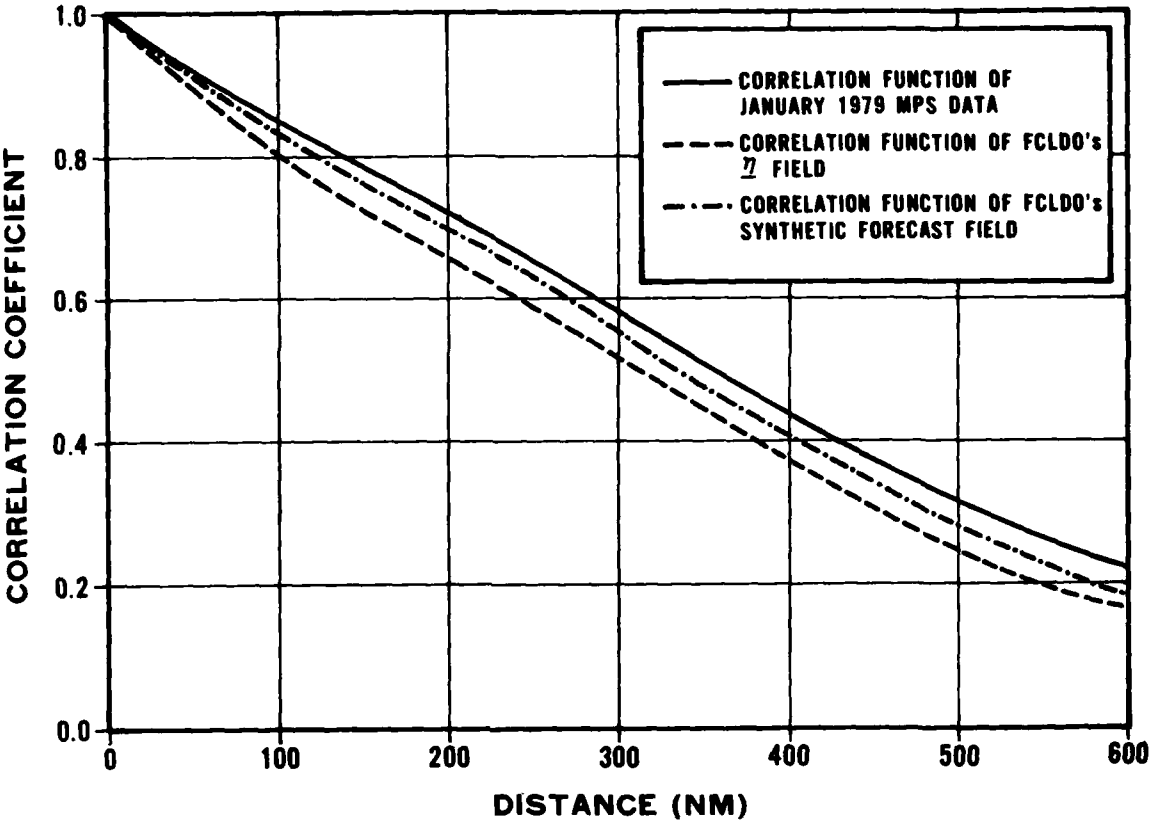


Figure 36. Comparison of the Spatial Correlation Functions of the Synthetic Forecast and $\underline{\eta}$ Fields Produced by FCLDO, with a Function Derived from January 1979 MPS Data. Input wavelengths of 2012.5 and 5175.0 km were used in the sawtooth wave generator. The input wavelengths were empirically derived to try to match a desired Gringorten curve for a scale distance of 11.5 km (after Boehm, 1977).

able, what happens if these cloud cover forecasts are more correlated or less correlated in space than the input MPS data? The spatial correlation function of the synthetic forecasts is determined by a combination of the correlation functions of: (a) the input observations, and (b) the $\underline{\eta}$ field of the model. The model user has no control over the spatial correlation of the input observations. That is a unique function of the input data set. However, the spatial correlation function of the $\underline{\eta}$ field can be controlled. By adjusting the

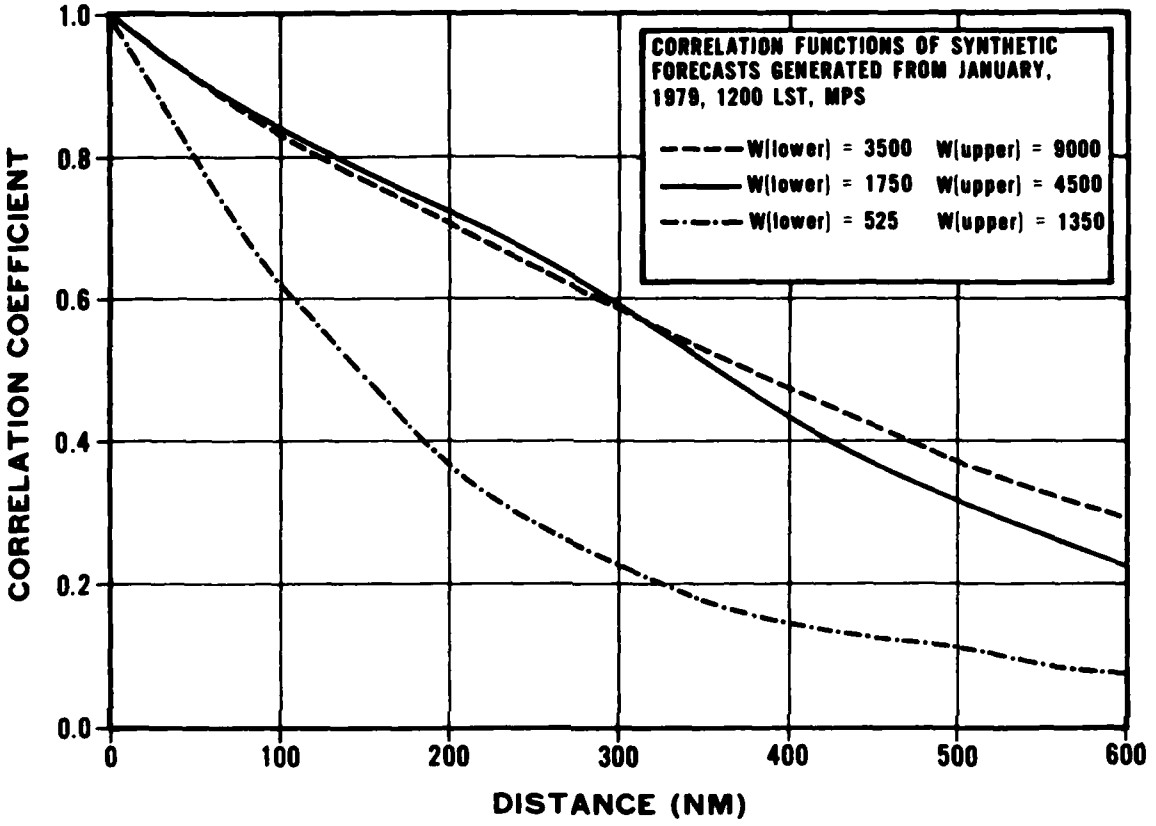


Figure 37. Comparison of the Spatial Correlation Functions of Synthetic Forecasts Produced by FCLDO. January 1979, 12 LST MPS data were used as input observation fields. Three different pairs of input wavelengths were used by the sawtooth wave generator. Curve ——— is virtually identical to the spatial correlation function of the input data.

input wavelengths for the model, the user can make the synthetic forecasts more or less correlated in space than the input observations. Figure 37 compares the spatial correlation functions of three sets of synthetic forecasts produced from the same input observations (January 1979, 12 LST, MPS data). For each set, ten forecasts were made from 31 different input observation fields. The only difference in the runs was the input wavelengths chosen for the sawtooth

wave generator. Note that the spatial correlation functions of the synthetic forecast fields can be changed significantly by selecting different pairs of input wavelengths.

An important aspect of fine-tuning the model was testing methods to conserve model run time. The sawtooth wave submodel is very CPU intensive. Processing time is measured in CPU hours for the longer simulation runs. One method that can be considered for reducing the amount of processing time is restricting the number of sawtooth waves emanating from each focal point.

Table 16. Comparison of CPU times for Various Runs of Cloud Forecast Simulation Model. A total of 310 forecast fields were generated by the model, on the 97 x 80 subgrid. The runs were performed on an IBM 4341, VM/370, DOX, using 3, 6, 9 and 12 sawtooth waves emanating from each focal point.

No. of Waves	Total CPU Time (sec)	Processing Time Saved (sec)	% CPU Time Saved
12	7566	---	---
9	6771	795	10.5
6	5078	2488	32.9
3	4093	3473	45.9

Table 16 shows the results of a test in which four sets of forecasts were produced by FCLDO. Ten forecast fields were generated from 31 different observation fields using January 1979 MPS data as input. Table 16 shows that the CPU time was reduced significantly as the number of waves was reduced. The question is, did these changes affect the shapes of the various spatial correlation functions? The spatial correlation function of the input observations obviously would not be affected by these changes and will not be discussed. Figure 38 compares the correlation functions of the η fields produced by the sawtooth

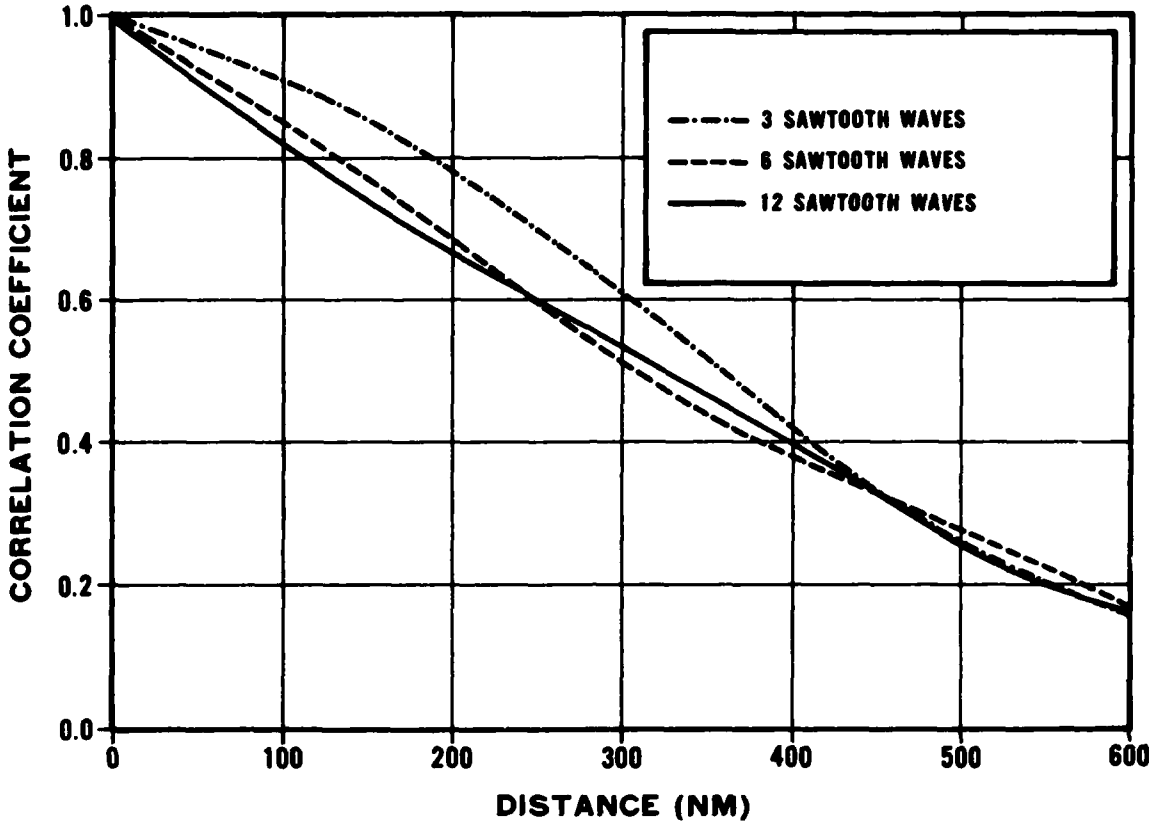


Figure 38. Comparison of the Spatial Correlation Functions of $\underline{\eta}$ Fields Produced by FCLDO. Input wavelengths were W(lower)=2200 and W(upper)=5500. A total of 310 fields were generated with 3, 6, and 12 sawtooth waves emanating from each focal point.

wave submodel, for 3, 6, and 12 sawtooth waves. There is virtually no difference in the functions for 12 and 6 waves. When only three waves emanate from each focal point, the shape of the spatial correlation function does change slightly. The $\underline{\eta}$ fields for three waves display slightly higher correlations in the first 300 NM and slightly lower correlations beyond 300 NM, than the other two cases.

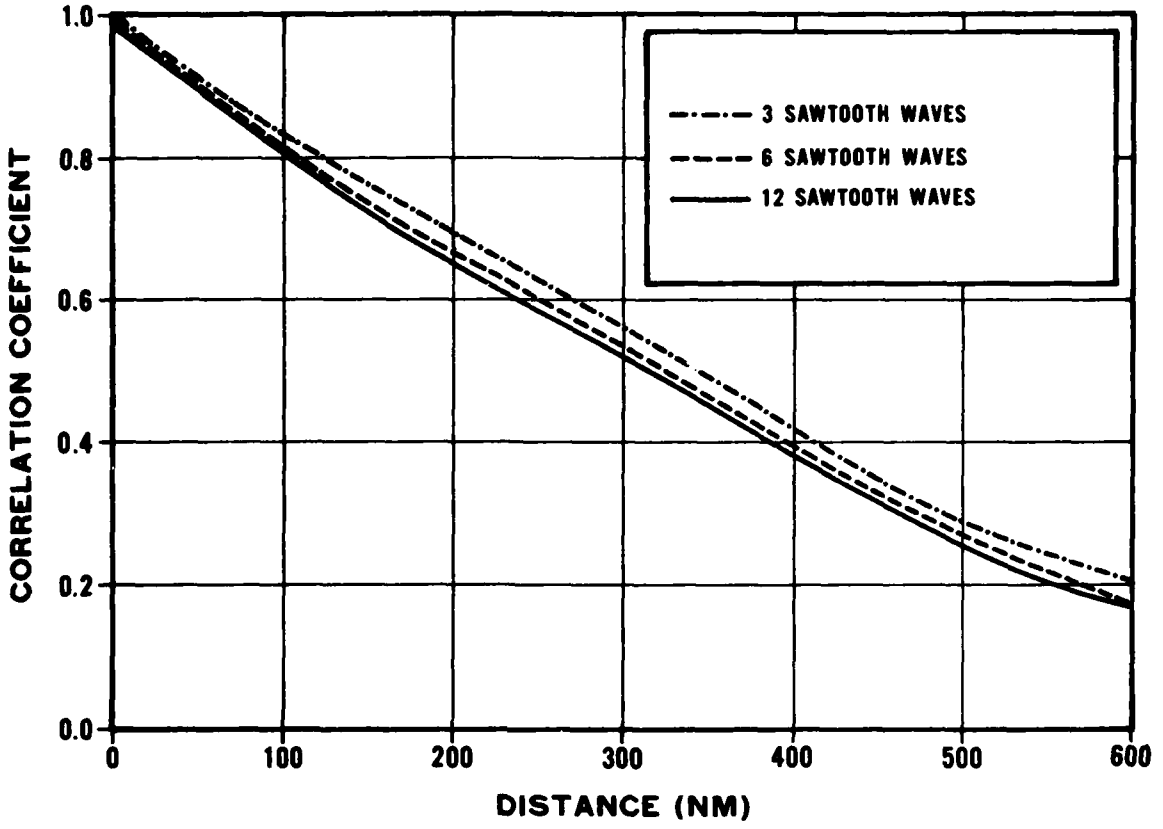


Figure 39. Comparison of the Spatial Correlation Functions of Synthetic Forecast Fields Produced by FCLDO. January 1979, 12 LST MPS data were used as input observation fields. Input wavelengths were W(lower)=2200 and W(upper)= 5500. A total of 310 forecast fields were generated with 3, 6, and 12 sawtooth waves emanating from each focal point.

Figure 39 illustrates again the fact that due to the short term of the synthetic forecasts for these particular tests, where forecast-observation correlation is high, the spatial correlation of the input observation fields exerts a greater influence on the forecasts than that of the $\underline{\eta}$ fields. Thus, the spatial correlation functions of the synthetic forecast fields for all three cases are virtually identical. A decision on selecting the number of waves for

the sawtooth wave generator should not be based on this single example. Factors, other than shorter computer run time, come into consideration. Foremost is the fact that tests of the model revealed that using too few sawtooth waves caused spurious interference patterns in the synthetic forecast fields. Additionally, for longer term forecasts, or for lower forecast-observation correlations, the synthetic forecasts for the three-wave case could be affected in the same manner as the $\underline{\eta}$ field for the three wave case. Based on model tests, it is recommended that no fewer than six sawtooth waves should emanate from each focal point in an operational model.

Chapter 6

MODEL ASSUMPTIONS AND LIMITATIONS

6.1 Introduction.

Models are generalizations or simplifications of complex reality. In order to construct such models, one usually must make certain simplifying assumptions that impose one or more limitations on the fidelity with which the model reproduces the complexities of reality. These limitations can be important in assessing the model's applicability in solving specific operational problems. The major assumptions and limitations of the Cloud Forecast Simulation Model are described in this chapter.

6.2 Basic Mathematical Assumptions in the Original Model.

A fundamental assumption made in the basic model was that the Johnson S_B curves, selected as the normalizing transformation function, could describe the cumulative probability distributions of observed and forecast cloud cover perfectly. Using RMS difference as an indication of closeness of fit, it was found that the marginal probability distributions from the hemispheric skill matrices could ordinarily be fit by the Johnson curves to within 2 percent. Maximum differences were usually less than 5 percent. This accuracy is usually adequate for most users' needs. Based on USAFETAC's experience in fitting over 300 cloud cover distributions to the Johnson curves, the closeness of fit is worst when trying to model relative frequency distributions that have multiple

relative maxima or minima, as demonstrated in Figure 40. In this case the RMS error was 4.9 percent, with a maximum difference of 11 percent between the observed and modeled distribution. Most users could not tolerate an error this large. Fortunately the example in Figure 40 was taken from an earlier USAFETAC project. Although the marginal probability distributions that were fit for the Cloud Forecast Simulation Project exhibited low RMS errors, the main problem was the modeled curves were biased toward the cloudy end.

Another key assumption in the basic model, and ultimately the reason the original technique was not selected for the operational model, is that the skill matrices to be preserved represent an underlying bivariate normal distribution of forecasts and verifying observations. As it turned out, the AFGWC SAVDOX skill matrices were not exactly bivariate normally distributed, and when coupled with the fact that the Johnson S_B distributions for the forecasts and observations were often biased toward the cloudy end, the result was that the basic model consistently did not adequately reproduce the input "desired" skill matrix.

As stated in Chapter 2, preserving the joint forecast-observation distributions described by the input skill matrices is a firm requirement of the model, one that could not be met using the basic technique. The final model eliminated the problems caused by these two assumptions by using the input skill matrix as the inverse-normalizing transformation (see Section 3.6 of this report). The forecast adjustment technique in turn imposed other limitations on the model that will be covered in Section 6.4, but the end result was that the input skill matrix could be reproduced without adversely affecting the other requirements imposed upon the model.

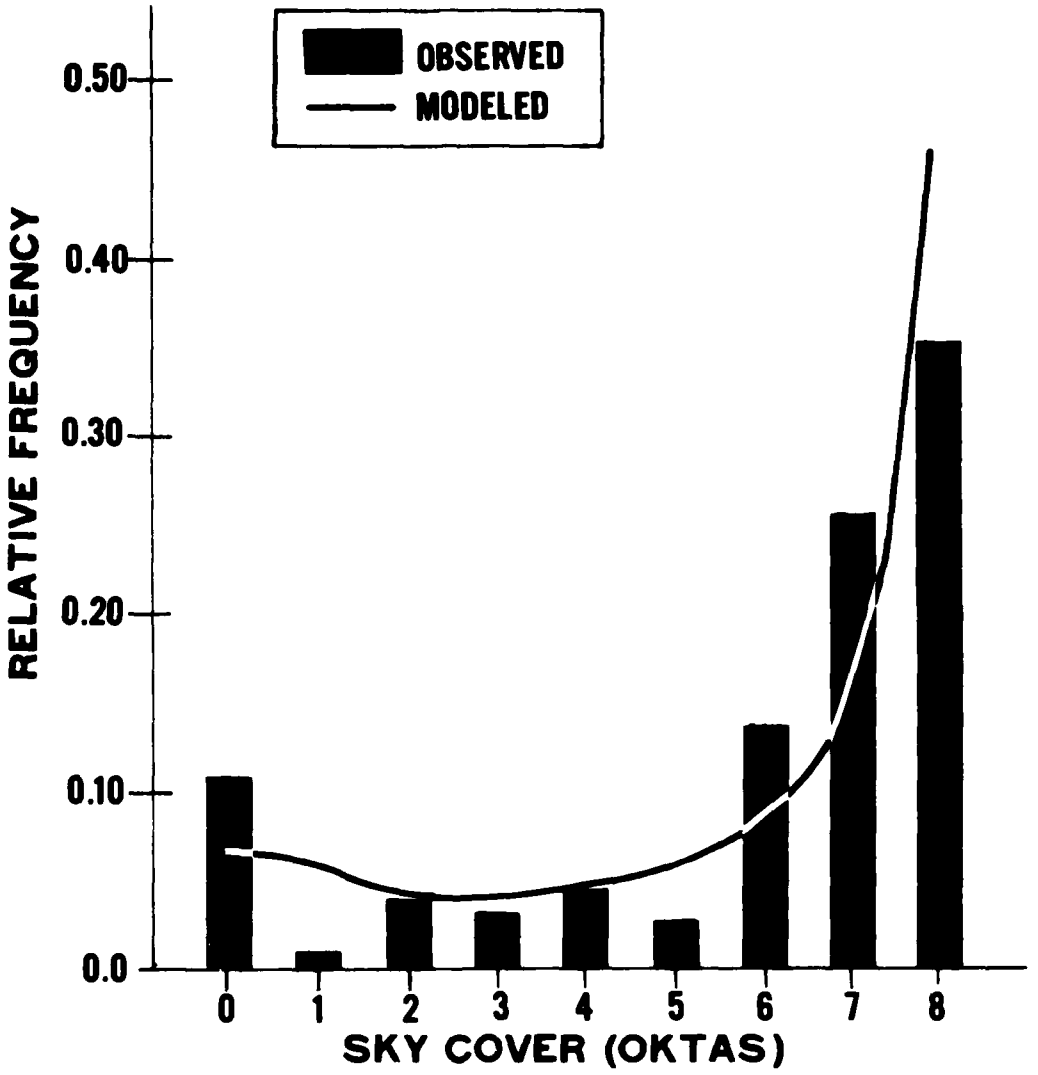


Figure 40. Example of the Type of Cloud Cover Distribution that the Johnson S_B Curve does not Handle Well. The relative frequency distribution of sky cover at Moscow, RS, for April, 15 Local Standard Time, and the Johnson S_B curve fit to that distribution are shown. The RMS between the observed distribution and modeled CDF is 3.7 percent, and the maximum difference is 11 percent.

6.3 Assumptions Dealing With Spatial Correlation.

Because preserving spatial correlation of the synthetic forecast fields was a particularly important requirement imposed upon the model, the assumptions dealing with spatial correlation are considered in some detail here.

An impediment in trying to simulate total cloud cover forecast fields realistically is that the actual forecast fields are not stored or archived. Shortly after verifying the AFGWC cloud prognosis model, AFGWC discards the forecast fields. No statistics on the spatial correlation of these forecasts are available for reference. This problem is discussed in more detail in Chapter 5. Briefly, the simulated forecasts are only 3 hours removed from the smoothed 3DNEPH (MPS) fields that are used to generate these forecasts, so it is assumed that the spatial correlation functions of these two types of fields should be similar.

The sawtooth wave submodel was tested and then modified so the spatial correlation function of the resultant random normal number field resembled that of Gringorten's Model-B for any given scale distance. The question is then, does Gringorten's Model-B adequately describe the types of spatial correlation functions found in real data fields? As with any model, the answer is that Gringorten's spatial correlation function cannot handle all types of observed functions. Although Gringorten's Model-B does do well most of the time, its main weakness is that it tends to approach zero in shorter distances than spatial correlation functions derived from real data.

The sawtooth wave submodel produces a random normal number field with an isotropic spatial correlation function. In such a function, spatial correlation is a simple function of distance alone, and does not depend on direction, i.e.,

$$\rho = f(\text{distance}), \qquad \underline{\text{not}} \qquad \rho = f(\text{distance, direction})$$

One might assume at first that the spatial correlation of meteorological variables is directionally dependent. Part of this project's study of smoothed 3DNEPH total sky cover (described in Chaper 5) included testing for directional dependence. This portion of the study involved calculating a correlation coefficient for every grid point within the 3DNEPH box with respect to the center point of the box. These fields were then hand analyzed. Figure 41 gives an example of the analysis performed. In each case the isopleths of equal correlation were nearly concentric circles around the center point. This result indicates that the isotropic assumption is acceptable for smoothed 3DNEPH total cloud cover and probably for the SAVDOX forecast fields as well.

The mathematical nature of the sawtooth wave submodel does not allow for any variability of the spatial correlation function over different portions of the random normal number field. This does not present a problem if the grid for the synthetic forecasts covers a limited area, say one 3DNEPH box. However, it is doubtful that many meteorological variables would display so homogeneous a spatial correlation function over a much larger area, such as a hemisphere or the whole globe. In the current application of the model, the synthetic forecasts are only 3 hours removed from the observation field, so the spatial correlation function of the observations will weigh quite heavily in

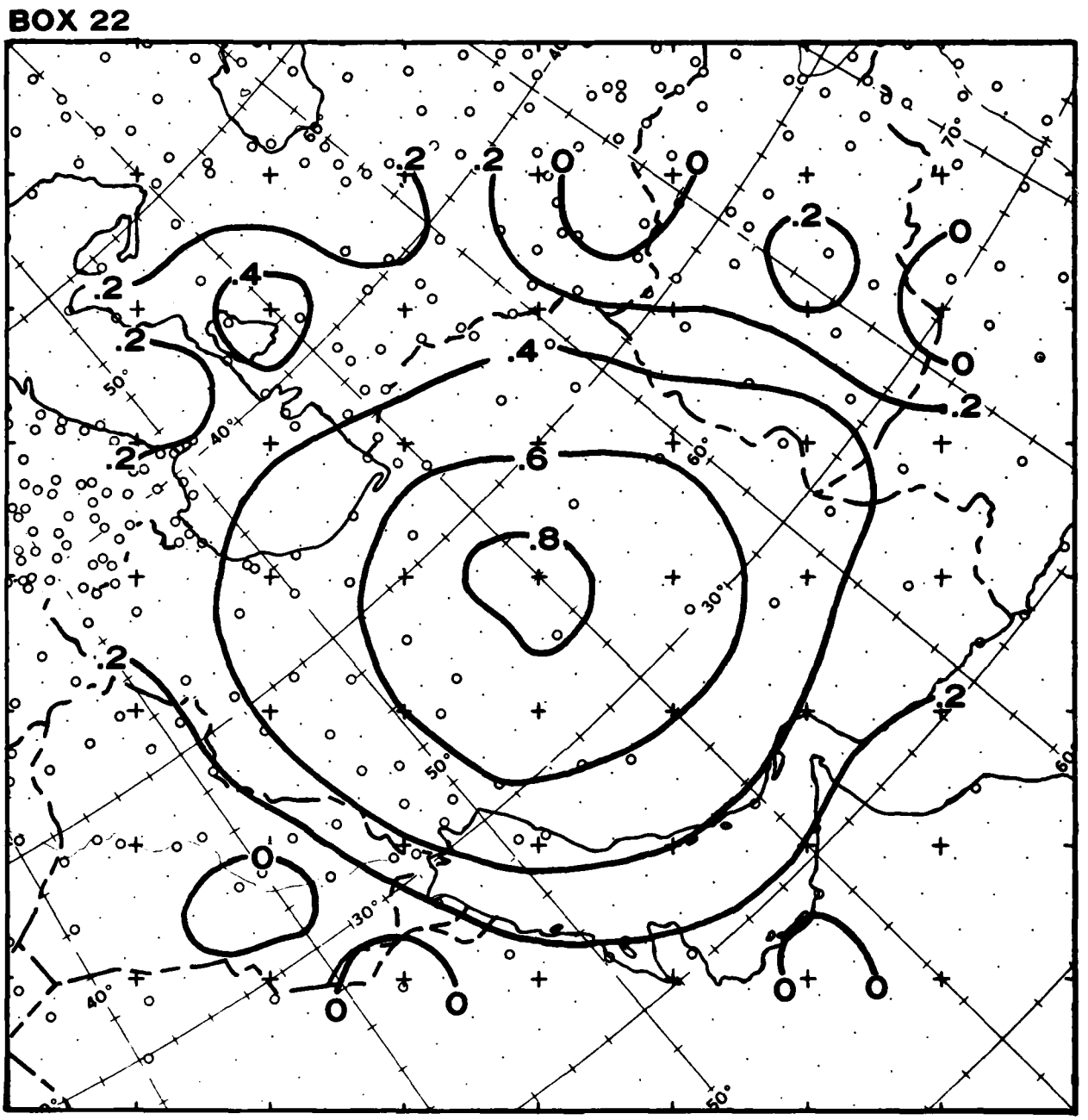


Figure 41. Spatial Correlation of Smoothed 3DNEPH Total Cloud Cover for Box 22, During April. Correlation coefficients were calculated for all grid points of the 3DNEPH box with respect to the center point.

determining the spatial correlation of the synthetic forecasts. This heavy weighting ascribed to the input observation fields in part allows the output synthetic forecast fields to display a geographical variability in their spatial correlation that the random normal number fields do not.

6.4 Limitations Imposed by the Input Skill Matrices.

Adjusting the synthetic forecasts to conform to the skill matrices, as is done in the final model, introduces some very real biases into the types of forecast fields produced by the model. The most significant error sources are sampling error and unrepresentativeness introduced by using a single input skill matrix for a large area.

6.4.1 Sampling Error. The forecast adjustment technique uses the forecast skill described in the input 21 x 21 matrix as an indication of an expected future forecast skill. This is the method of inferential statistics, in which one samples from a population in order to infer certain elements of the character of that population. Such inferences are made under uncertainty because no finite sample provides a complete description of the population. The uncertainty inherent in inferring population parameters from sample statistics depends on the sampling methods used. In the case of the 21 x 21 skill matrices, AFGWC is sampling preferentially, not randomly. Remembering that the skill matrices are built as a verification tool to test the accuracy of the operational AFGWC cloud prognoses, it must also be kept in mind that the verification points are chosen nonrandomly and nonuniformly over space and time. For example, verification points are almost never located over oceanic areas.

The skill matrix for a particular month should not be expected to describe perfectly the forecast skill for all such months in the future.

6.4.2 A Single Skill Matrix vs. Regional Skill Matrices. The forecast distributions derived from the input skill matrices are assumed to describe the performance of AFGWC cloud forecast model at each forecast point of the simulation model's grid, when in reality they do not. When the model is operating on a hemispheric scale, inconsistencies can arise from using one hemispheric skill matrix as the basis for all simulated forecasts anywhere in the hemisphere. This can, in part, be overcome by using several "regional" skill matrices when the model is producing synthetic forecast fields on the hemispheric or global scale. There will be some problems involved in developing regional skill matrices because some areas of the globe are sampled repeatedly with great density, while others are barely sampled at all. The resultant matrices from the sparsely sampled areas may be quite pathological in nature and may not be good descriptors of forecast skill in their area; yet the final simulation model will reproduce these matrices perfectly, however unrepresentative they may be. Another problem might be inconsistencies at boundaries between areas with different skill matrices. Despite the possible problems involved in incorporating regional skill matrices into the large-scale models, the fact that the model will be producing fields that are more consistent with the operational AFGWC cloud prognoses should make this simulation technique more appealing to potential customers. Otherwise the final product will be a spatially correlated and statistically sound forecast field, but one that might not be very realistic for any particular location on the globe.

Chapter 7

CONCLUSIONS AND RECOMMENDATIONS FOR OPERATIONAL IMPLEMENTATION

The basic design goal of the Cloud Forecast Simulation Model is to generate "realistic" synthetic total cloud cover forecast fields valid at time t from input observed total cloud cover fields, also valid at time t. The model does this by generating spatially correlated, random equivalent normal deviate (END) fields, and then using these fields to decorrelate the synthetic forecast fields from the input observed fields. For these forecast fields to be considered "realistic", they must be statistically consistent with the set of operational SAVDOX total cloud cover forecast fields produced by AFGWC. Specifically, they should have the same skill (as defined by a set of SAVDOX skill matrices) and spatial corrrelation as the operational product.

The final version of the Cloud Forecast Simulation Model, FCLDO, meets these requirements and produces forecast fields that are "realistic" by the definition above. The user should be aware of the model's assumptions and limitations, listed in Chapter 6, which still distinguish the set of synthetic forecasts from the operational product. Despite these differences, the limitations of the Cloud Forecast Simulation Model are very minor when compared with those of the earlier, uncorrelated forecast simulation technique (see Chapter 1), and the model's ability to produce spatially coherent synthetic forecast fields represents a substantial improvement over that earlier method.

The hybrid sawtooth wave/skill matrix design used in the final version of the Cloud Forecast Simulation Model made achieving the customer's goal of exact reproduction of the AFGWC skill matrices possible; however, it probably has only limited usefulness beyond this particular application. This is because it uses categorical skill matrices describing forecast performance directly in generating its synthetic forecast fields. In addressing the class problem of producing synthetic forecast fields for a wide range of meteorological variables, where such skill matrices are seldom available, the technique is impossible to apply. In the early version of the Cloud Forecast Simulation Model, however, the input variable describing forecast skill, the forecast-observation correlation, can be modeled as a function of time using equations such as Equation (3). Thus, if a Gringorten Model-B scale distance describing the spatial correlation of the observed fields and suitable normalizing transforms for forecasts and observations can be found, the basic Cloud Forecast Simulation Model, with minor modifications, could produce synthetic two-dimensional forecast fields for almost any continuous meteorological variable.

The development of this two-dimensional forecast field simulation capability lays much of the groundwork for the solution of a similar class problem, the simulation of two-dimensional observed fields. In fact, given that a suitable normalizing transform for the desired observed variable can be found, the solution is merely to use the sawtooth wave submodel to produce random END fields with spatial correlations similar to the desired observed fields, and then to inverse-normalize, yielding the desired synthetic fields of observations.

The Cloud Forecast Simulation Model was produced as a technique development effort rather than a computer software development project. To permit testing of the model, an experimental, demonstrational computer program called FCLDO (Cloud Forecast Simulation Model, Level 0) was written, and various versions of it were made available to the customer during the course of the project. This program was intended primarily to be demonstrational in nature and allows for optional running of either the original (basic) or final (hybrid) versions of the model. The computer program was written in simplistic FORTRAN so as to run on almost any computer with only minor modification. Extensive documentation was included within the program to facilitate modification by the customer to fit his operational simulation requirements.

No effort was made in developing the demonstrational Cloud Forecast Simulation Model to conserve either core storage or run time required by the program. Conversely, every effort was made to include all possible options and diagnostics, with the intent that the customer could remove those portions he felt were superfluous. In this vein, several actions may be taken to create a more efficient operational version from the FCLDO program. First, most of the statistical accumulating and checking code can be removed, leaving only as many diagnostics as the user feels are necessary. Second, if the user has categorical observations and categorical skill matrices available, and desires only categorical forecasts, the normalized observed and forecast cloud cover arrays and much of the normalizing code can also be disposed of. And third, to conserve run time, the number of sawtooth waves used by the model may be reduced to a minimum of six waves. If all these actions are taken, the model's run time and core storage requirements can both be reduced significantly. It is further suggested that if using categorical skill matrices, the geographical

area which contributes to building each matrix should be small enough so that the cloud cover (or other parameter) climatology is essentially homogeneous over the region, but not so small that the skill matrices suffer unacceptable sampling error. This should help to minimize the final Cloud Forecast Simulation Model's limitation of producing identical forecast distributions at every forecast point.

REFERENCES

Acton, F. S., 1970: Numerical Methods that Work. Harper & Row, Publishers, New York, pp 41-88.

Boehm, A. R., 1976: Transnormalized Regression Probability. Air Weather Service, AWS-TR-75-259, 52 pp.

Boehm, A. R., 1979: Project Notes for COLOSSUS Weather-A Model. Unpublished Manuscript, USAFETAC.

Gringorten, I. I., 1979: Probability Models of Weather Conditions Occupying a Line or Area. J. Appl. Meteorol., 18, pp 957-977.

Naylor, T. H J., et al, 1966: Computer Simulation Techniques. John Wiley & Sons, Inc., New York, p 93.

Somerville, P. N., and S. J. Bean, 1979: A New Model for Sky Cover. University of Central Florida, Department of Mathematics and Statistics, Scientific Report No. 5, AFGL-TR-79-0219, 37 pp.

Somerville, P. N., S. Watkins, and R. Daley, 1978: Some Models for Sky Cover. Florida Technological University, Department of Mathematics and Statistics, Scientific Report No. 2, AFGL-TR-78-0219, 26 pp.

Panofsky, H. A., and G. W. Brier, 1958: Some Applications of Statistics to Meteorology. The Pennsylvania State University, University Park, Pennsylvania, pp 100-104.

Appendix A

CALCULATION OF SPATIAL CORRELATION OF 3DNEPH DATA

a. In order to estimate the spatial correlation functions of smoothed SAVDOX fields, an extensive study of smoothed 3DNEPH data was undertaken. The spatial correlation functions of these two data types were assumed to be similar. The study also provided information on the effects of the smoother and helped verify whether the Cloud Forecast Simulation model was producing realistic results.

b. The basic process of the Cloud Forecast Simulation Model is to transform input sky cover observations separately into equivalent normal deviates (ENDs), generate an END of a forecast based on the input observation using the forecast simulation equation, and then transform the ENDs for the forecasts into cloud cover amounts. This involves assuming that these individually normalized variables are distributed multivariate normally.

c. To remain consistent with the basic assumption of multivariate normal distribution in the simulation, the spatial correlation of ENDs of 3DNEPH total sky cover was studied rather than the correlation of the raw 3DNEPH sky cover data. Two of the most frequently used methods for calculating correlation coefficients are (1) the Pearson product moment formula, and (2) the tetrachoric method. The Pearson product moment formula is defined by the equation,

$$r_{xy} = \frac{\overline{XY} - \bar{X}\,\bar{Y}}{\sqrt{\overline{X}^2 - (\bar{X})^2}\ \sqrt{\overline{Y}^2 - (\bar{Y})^2}} \qquad \text{(A-1)}$$

Although it is relatively easy to implement on a computer, Pearson's method has the inherent disadvantage of requiring the raw data to be available for the computations. To use Pearson's method on the ENDs of the 3DNEPH data would involve the time consuming and cumbersome procedure of processing the 3DNEPH data tapes twice: once to determine probability distributions of the 3DNEPH data, and again to transform each observation to an END and then perform the correlation computations. The tetrachoric method offers the advantage of being able to use data that have already been categorized. By using this method, the 3DNEPH tapes had to be processed only once, saving valuable CPU time. During processing, the data were reduced to a series of two-by-two contingency tables.

d. Any two variables can be reduced to a two-by-two table:

Variable 1

Variable 2 (V_{t2})	Above (V_{t1})	Below
Above	A	B
Below	C	D

where A, B, C, and D are the number of cases above or below the critical, or threshold, values (V_t) of the respective variables. An approximation to the tetrachoric correlation coefficient (r_t) can be obtained by Equation (A-2):

$$r_t = \sin\left(\frac{\pi}{2}\,\frac{\sqrt{AD} - \sqrt{BC}}{\sqrt{AD} + \sqrt{BC}}\right) \qquad \text{(A-2)}$$

This equation is accurate when (A+B)/N and (A+C)/N are close to 0.5 (where N is the total number of cases) but may contain sizeable errors for values near one or zero. Since there is no simple exact formula for calculating r_t, an algorithm based on the false position method (Acton, 1970) is used by USAFETAC. The coefficient is evaluated at two initial guess values, and linear interpolation is used to find a better estimate. The quantity r_t behaves in a manner similar to an ordinary linear correlation coefficient, but the exact numerical value is not completely comparable. The value of r_t varies from -1 to +1, giving zero for no relation, but the sign depends in a rather arbitrary manner on the arrangement of the contingency table.

e. As an illustration of these three methods, let us turn to an example using smoothed 3DNEPH data for box 23, January. The goal is to compare the spatial correlation coefficients calculated by these three methods for a grid point 100 NM from the center of the box. In all, 248 pairs of observations were used for this example. X will represent the ENDs of cloud cover over the center point, Y will represent the ENDs of cloud cover over the reference point.

(1) The statistical values for X and Y obtained from the observations were,

$$\overline{X} = -0.035 \qquad \overline{Y} = 0.127$$

$$\overline{X^2} = 0.794 \qquad \overline{Y^2} = 0.739$$

$$\overline{XY} = 0.508$$

Substituting these values into Equation (A-1), the Pearson product moment correlation coefficient is,

$$r = \frac{0.508 - (-0.035)(0.127)}{\sqrt{0.794 - 0.001}\ \sqrt{0.739 - 0.016}} = 0.718$$

(2) The two-by-two table that resulted was:

		X: Above (X_t=0)	X: Below
Y: ABV (Y_t=0)		82	28
Y: BLW		33	105

using Equation (A-2), the sine approximation for the tetrachoric correlation coefficient becomes:

$$r_t = \sin\left(\frac{\pi}{2}\,\frac{\sqrt{82.105} - \sqrt{28.33}}{\sqrt{82.105} + \sqrt{28.33}}\right) = 0.714$$

(3) Finally, by feeding the two-by-two table into the USAFETAC subroutine TETRA, which calculates tetrachoric correlation iteratively, a better estimate of $r_t = 0.712$ is obtained.

f. 3DNEPH data tapes are arranged by box, month, and year. For example one tape contains all data for box 22, January 1973. Six to eight tapes may be required to obtain all data for a specific box and month. For this spatial correlation study, daily fields of total cloud cover at 12 LST were extracted. The resulting 1/8-mesh 64 x 64 array was reduced to a 31 x 31 1/4-mesh array by applying a nine-point 4-2-1 smoother (see Figure 21 in the main text for an example of the smoother). Two-by-two tables were constructed using the center point value and values of points at various distances as data pairs. Specifically, tables were built in each grid system for four points at distances of 50 and 100 NM, and for 12 points at distances of 150 to 600 NM, by 50-NM increments. The position of these grid point pairs are shown in Figure A-1. The two range marks represent the 100- and 200-NM distance circles superimposed on the center of a 3DNEPH box. The grid points shown are spaced 50 NM (1/4-mesh). The 3DNEPH data are at 1/8-mesh resolution, but on this figure the intermediate points are omitted. For distances less than or equal to 100 NM the two-by-two tables were constructed for all pairs in the same relative positon as the C-A grid pairs in the figure. For distances 150 to 600 NM, additional points were

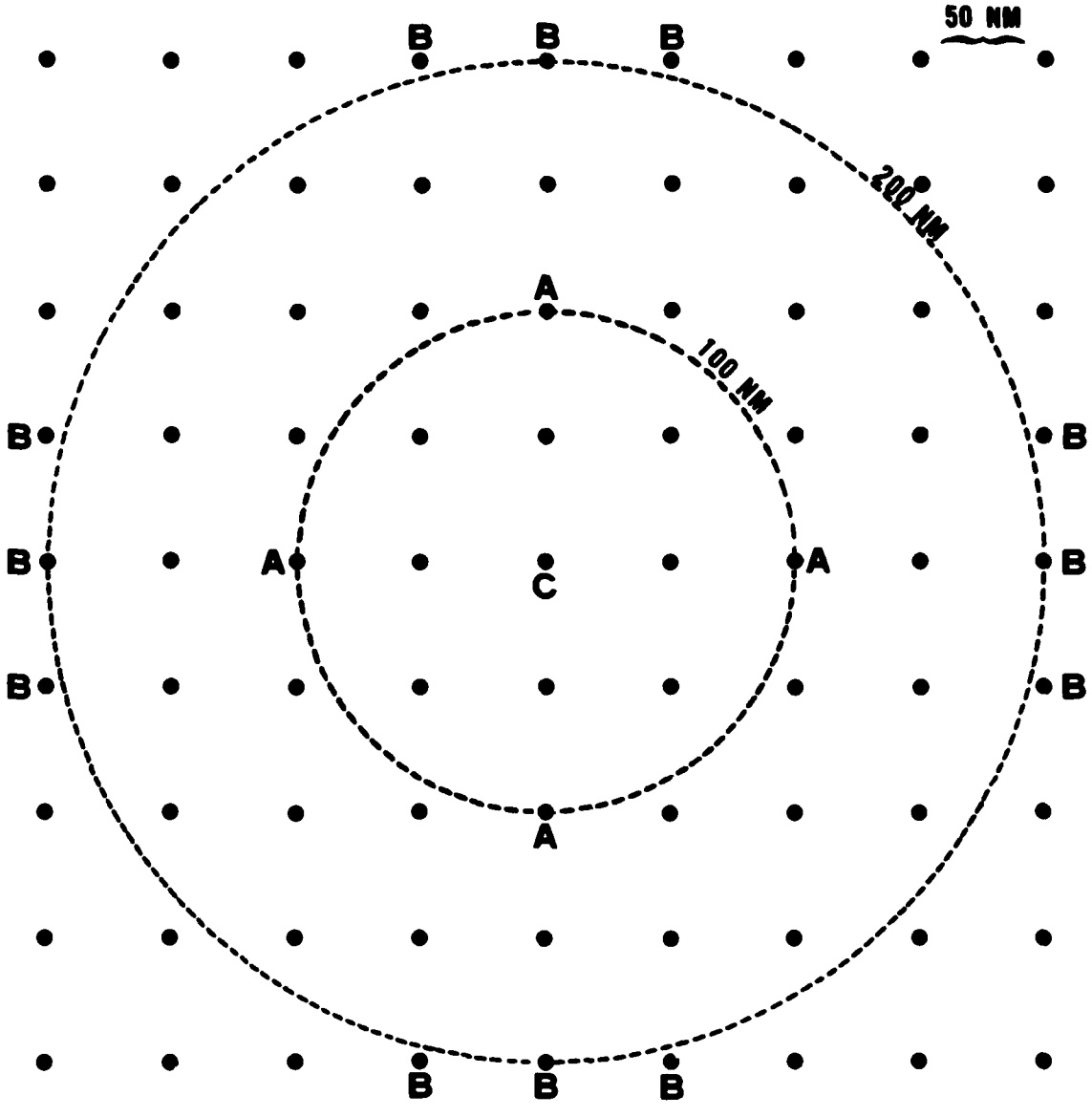


Figure A-1. Quarter-mesh Grid Showing Location of Data Pairs Used in the Tetrachoric Correlation Calculations.

added, and tables were built for all pairs that are in the same relative positon as the C-B grid pairs in the figure. This process was continued until data for all available years were accumulated. Correlation coefficients were then calculated from the contingency tables, and average correlation coefficients at

the various distances were used to fit to an appropriate Gringorten Model-B curve. In this manner characteristic scale distances for both smoothed and raw 3DNEPH data was acquired over thirteen different 3DNEPH boxes of the Northern Hemisphere.

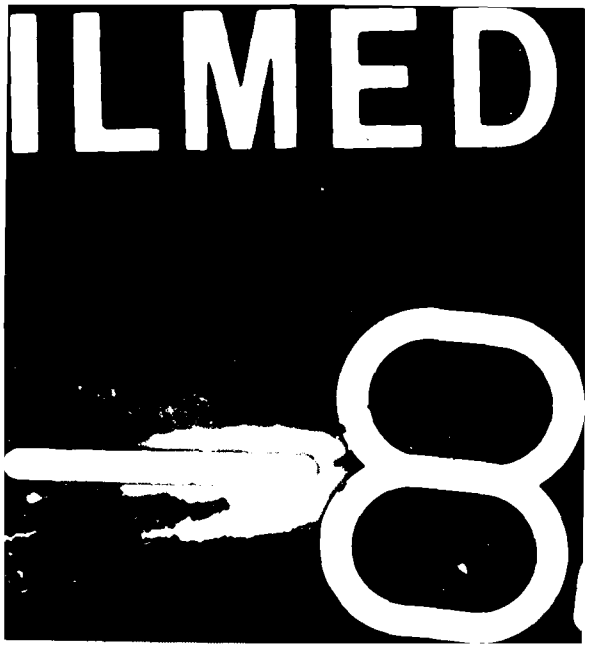
LIST OF ABBREVIATIONS AND ACRONYMS

3DNEPH	Three-dimensional Nephanalysis, an AFGWC/USAFETAC Cloud Data Set
5LYR	AFGWC Five-layer Cloud Prognosis Model
AFGWC	Air Force Global Weather Central
AWS	Air Weather Service, a Technical Service of MAC
CPU	Central Processing Unit
DOS	IBM Disk Operating System
DOX	IBM Disk Operating System Enhanced by SDI GRASP
END	Equivalent Normal Deviate, or Standard Normal Variable
FCLDO	Cloud Forecast Simulation Model, Level 0
FCLDJ	Johnson S_B Curve Fitting Program
GMT	Greenwich Mean Time, or "Zulu" (Z) Time
GRASP	IBM DOS Enhancement Offered by SDI
HRCP	AFGWC High-resolution Cloud Prognosis Model
IBM	International Business Machines Corporation
km	Kilometers
LST	Local Sun Time
MAC	Miliary Airlift Command, a Specified USAF Command
MPS	Multi-purpose Simulator, Cloud Cover Data Set
NM	Nautical Miles
RMS	Root Mean Square
RMSE	Root Mean Square Error
SAVDOX	AFGWC Cloud Forecast Data Set
SB	Johnson Single-bounded Probability Distribution
SD	Standard Deviation
SDI	A Software Design Firm Offering GRASP
SM	Statute Miles
TALON	Tactical Air/Land Operations Simulator
TRONEW	AFGWC New Tropical Model, a Cloud Prognosis Model
USAF	United States Air Force
USAFETAC	U.S. Air Force Environmental Technical Applications Center
VM	Virtual Machine
VM/370	Virtual Machine/System 370

DISTRIBUTION:

HQ AWS, AWS Wing HQ, AWS Squadron HQ, Selected Detachments and OLs, plus Special Distribution

U.S. GOVERNMENT PRINTING OFFICE: 1981—565-305/125


ILMED